Kaye Stacey · Ross Turner *Editors*

Assessing Mathematical Literacy

The PISA Experience

Springer

Assessing Mathematical Literacy

Kaye Stacey • Ross Turner
Editors

Assessing Mathematical Literacy

The PISA Experience

Springer

Editors
Kaye Stacey
Melbourne Graduate School of Education
The University of Melbourne
Melbourne, VIC, Australia

Ross Turner
International Surveys, Educational Monitoring and Research
Australian Council for Educational Research (ACER)
Camberwell, VIC, Australia

ISBN 978-3-319-10120-0 ISBN 978-3-319-10121-7 (eBook)
DOI 10.1007/978-3-319-10121-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014954104

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The initiative for this book came from the PISA 2012 Mathematics Expert Group, which had worked together with a team from the Australian Council for Educational Research (ACER) for nearly 4 years in the preparation of the OECD's 2012 PISA survey. The mathematics assessment for the 2012 survey underwent substantial changes, building on and further developing the structures and conceptualisation of the 2003 survey (when Mathematics had last been the major domain) and responding to the wide-ranging international feedback that had arisen in those 9 years. The Framework has grown steadily since its inception for the 2000 survey, and its impact has expanded dramatically over this time. The item design has also been substantially refined. The expert group came to realise that the work that goes into an international survey such as PISA should be better known: hence this book. We hope it is a contribution both to thinking about the most fundamental goals and activities of mathematics education and toward better understanding the results of the PISA surveys.

It has been a pleasure to work with a team of such talented, engaged, and well-informed authors in the preparation of this book. Many chapter authors were also members of the Mathematics Expert Group for the PISA 2012 survey and the mathematics teams of international contractors for PISA 2012 led by ACER. We thank them for contributions to the book as well as for their contribution to the Mathematics Framework and items for the 2012 survey. Other authors have played important roles in using PISA to improve mathematics education in their own countries. The editors have also enjoyed bringing their own two different perspectives together as they worked on this book: Ross's experience as the leader of the ACER team responsible for delivering the mathematics framework, items, and coding since the first PISA survey and Kaye's view from research, teaching, and national policy and as chair of the Mathematics Expert Group for PISA 2012.

It is essential to acknowledge that many of the ideas in the book are the outcome of the joint work of the members of all the Mathematics Expert Groups from PISA 2000 to PISA 2012. Their names are listed at the end of this Preface along with

other key mathematics staff members of agencies contracted to develop and implement PISA mathematics over its first several survey administrations.

We also acknowledge the valuable input of the Springer editors and especially of the anonymous reviewers whose useful comments helped sharpen the text. It is a special pleasure to acknowledge the work of Pam Firth from the University of Melbourne for her able editorial and administrative assistance.

Opinions expressed in this book are those of the authors and do not imply any endorsement by the Organisation for Economic Co-operation and Development (OECD) or any other organization.

Melbourne, VIC, Australia — Kaye Stacey
Camberwell, VIC, Australia — Ross Turner
3 Dec 2013

Membership of Mathematics Expert Groups and Other Contributors

2000

Jan de Lange (Chair, Netherlands), Raimondo Bolletta (Italy), Sean Close (Ireland), Maria Luisa Moreno (Spain), Mogens Niss (Denmark), Kyungmee Park (Korea), Thomas Romberg (United States), Peter Schüller (Austria)

Margaret Wu and Ross Turner (ACER, Executive Officers)

2003

Jan de Lange (Chair, Netherlands), Werner Blum (Germany), Vladimir Burjan (Slovak Republic), Sean Close (Ireland), John Dossey (United States), Mary Lindquist (United States), Zbigniew Marciniak (Poland), Mogens Niss (Denmark), Kyungmee Park (Korea), Luis Rico (Spain), Yoshinori Shimizu (Japan)

Ross Turner (Executive Officer)

2006

Jan de Lange (Chair, Netherlands), Werner Blum (Germany), John Dossey (United States), Zbigniew Marciniak (Poland), Mogens Niss (Denmark), Yoshinori Shimizu (Japan)

Ross Turner (Executive Officer)

2009

Jan de Lange (Chair, Netherlands), Werner Blum (Germany), John Dossey (United States), Zbigniew Marciniak (Poland), Mogens Niss (Denmark), Yoshinori Shimizu (Japan)

Ross Turner (Executive Officer)

2012

Kaye Stacey (Chair, Australia), Caroline Bardini (France, Australia), Werner Blum (Germany), Solomon Garfunkel (USA), Joan Ferrini-Mundy (USA), Toshikazu Ikeda (Japan), Zbigniew Marciniak (Poland), Mogens Niss (Denmark), Martin Ripley (England), William Schmidt (USA)

Ross Turner (Executive Officer)

Other Contributors

We acknowledge the contribution of other ACER staff members, consultants and staff of organisations working closely with ACER to develop PISA mathematics over its first several administrations, but who did not contribute directly to writing this book.

Kees Lagerwaard, Gerben van Lent (both formerly of Cito in the Netherlands), Hanako Senuma (formerly of the National Institute for Educational Policy Research, NIER, in Japan), Margaret Wu (formerly of ACER), Raymond J. Adams (ACER), Béatrice Halleux (HallStat SPRL, Belgium).

Contents

The Assessment of Mathematical Literacy: Introduction to PISA and to This Book

Abstract This book gives the 'inside story' of the mathematics component of the PISA survey, with contributions from authors directly involved in the international PISA development and implementation, and national policy responses and practical actions. This introductory chapter introduces the key ideas explored in the book, and sets the context in which detailed commentary is provided in later sections. The two main audiences for the book are identified as those with direct involvement or concern with what happens in mathematics classrooms (that is, mathematics teachers, curriculum developers, test developers and teacher educators), and those with an interest in the policy environment within which mathematics education occurs. The three main parts of the book are introduced. The first part describes the key concepts of the Mathematics Framework and their evolution over the PISA 2000 to PISA 2012 survey administrations, including the literacy concept, the place of mathematical modelling, and of mathematical competencies. The second part gives an insider view of the development and implementation of the PISA survey, including test item development and test administration, questionnaire development and the new computer-based assessment of mathematics. The third part gives a collection of reports and views about impacts of the PISA survey in 14 countries. This introductory chapter also gives a very broad outline of the PISA surveys for mathematics for readers unfamiliar with the details of this initiative.

Aims of This Book

This book aims to give the 'inside story' of the mathematics component of the world's largest educational survey—the assessment of mathematical literacy of students around the world by PISA, the Programme for International Student Assessment of the Organisation for Economic Co-operation and Development (OECD). The editors and authors have been directly involved in creating the PISA Mathematics Framework and mathematical literacy test items and designing and implementing the associated quality control measures. Some contributors have been through all first five administrations of the PISA survey: PISA 2000, PISA 2003, PISA 2006, PISA 2009 and PISA 2012. Other authors are also involved in understanding and interpreting the results of the PISA surveys in their own countries and in designing initiatives to improve their educational systems in response to PISA results. The initiative for the book came from the PISA 2012 Mathematics Expert Group, the members of which worked together with the team of international contractors led by The Australian Council for Educational Research (ACER) for nearly 4 years in the preparation of the Framework and items for the 2012 survey. The conduct of international assessments involves many groups: the commissioning governments, the psychometricians who ensure that the statistical basis of the survey makes the results sufficiently authoritative for legitimate comparisons to be made, the psychologists and educators who design the parameters and variables of interest across the whole study, and groups in each participating economy who work with schools and students and with the policy implications arising from the assessments. Within this large mix, the Mathematics Expert Group is the voice of mathematics educators, and this book looks at PISA mainly from their point of view.

All members of the Mathematics Expert Group felt strongly that the theoretical and practical developments of PISA needed to be better known. Naturally the main interest of PISA is in its results: the country achievements, rankings and trends over time, the examination of equity, the links between performance and characteristics of schools and teachers. However, this book is not about the results. Instead it has been written in the belief that the results of PISA will be used most wisely by people who understand what lies behind PISA, both in its conceptualisation and in the practical issues of designing and conducting a valid and equitable survey of a worthwhile construct.

The editors and authors had two particular sets of interests at the forefront of their thinking as the included material was selected and presented: those with direct involvement or interest in what happens in mathematics classrooms; and those with an interest in the policy environment within which mathematics education and testing occurs.

First, mathematics teachers, curriculum developers, test developers and teacher educators will be interested in the detailed discussion of the mathematical literacy concept, the processes of development of PISA mathematics tasks, the results of research into key drivers of mathematical literacy and the way that literacy is expressed in the behavioural repertoire of mathematics students; and the insiders' insights into the practical examples of mathematics tasks that have been used in the PISA surveys.

Second, the community of interests that has generated or supported the PISA survey will also be interested in several aspects of this book. Those responsible for guiding the development and implementation of PISA may enjoy their share of the credit for producing such a significant program that has been so influential in shaping educational practice in so many ways. One part of the book aims to document ways in which this has happened, from experts around the world. At an individual country level, those responsible for various aspects of educational policy development may benefit from the observations presented here regarding the specific ways in which PISA ideas and methodology have been, are being or could be used to drive educational improvement. There are many practical models to follow.

The book is in three main parts. Part I begins with a discussion of the concept of mathematical literacy. The origins of this concept are drawn together, along with some of the closely related and often partially conflicting ideas that sit alongside it. These are discussed to clarify the different terminology that has been used particularly in recent years to discuss this part of the mathematics education territory. The PISA Mathematics Framework is introduced as a significant milestone in the development and also in the dissemination of these ideas, because the survey is used so widely around the world (65 countries in 2012). The underlying mathematical competencies on which mathematical literacy so strongly depends are described in two chapters, along with a scheme for operationalising these competencies so that the cognitive demand of items can be estimated. PISA assesses 15-year-olds' ability to apply knowledge and skills to real-life problems rather than how well they have learned the school curriculum. For this reason, there is a chapter that focuses on the links in the assessment between the real world and the mathematical world. This first part concludes with a personal reflection from a research mathematician on how his views of mathematics education have changed as a result of his involvement with PISA mathematics. Although the value of education for all is now widely acknowledged, exactly what type of mathematics should be given the highest priority remains contested.

Part II provides significant detail on aspects of the development and implementation of the PISA survey, specifically the processes of mathematics item development in paper-based and computer-based environments, coding of responses to items, and questionnaire development. Some of the tricks of the trade used by one of the world's pre-eminent test development agencies are discussed; features and characteristics of several publicly released PISA items are demonstrated; and issues that affect the ways in which these mathematics items are used to measure levels of mathematical literacy are canvassed. This part also describes how the PISA 2012 survey collected data to measure the opportunity students in participating countries have had to learn mathematics involving the approaches promoted through PISA. Evidence from sources such as this can assist countries to find the right balance of PISA-like classroom activities with traditional approaches to mathematics, when the goal is mathematical literacy. A major theme of this part is the range of quality assurance measures that need to be applied so that the results of PISA are meaningful, and the substantial international collaboration that is involved in doing this complex task.

The third part of the book goes to the issue of impact. We present the viewpoints of mathematics educators in various contexts in 14 countries to show how PISA and its constituent ideas and methods have influenced teaching and learning practices, curriculum arrangements, assessment practices at a variety of levels, and the education debate more generally in different countries. Some of these contributions may go some way to explaining why there has been improvement in PISA scores in some countries over time and may provide models and ideas for policy makers who wish to use PISA outcomes as a stimulus for further educational improvement.

A Compact Introduction to PISA Surveys

This part gives a very brief introduction to the PISA program, designed to provide background information for readers unfamiliar with PISA and pointing to later sections of the book where particular issues mentioned are developed in greater detail.

PISA stands for the Programme for International Student Assessment, which was initiated by the Organisation for Economic Co-operation and Development (OECD) in the 1990s to provide governments and other interested parties information on the effectiveness of educational systems, especially in preparing students for the challenges of their future lives. The foreword to the first report of results from PISA sets out the agenda in these terms:

> PISA represents a new commitment by the governments of OECD countries to monitor the outcomes of educational systems in terms of student achievement on a regular basis and within a common framework that is internationally agreed upon. PISA aims at providing a new basis for policy dialogue and for collaboration in defining and operationalizing educational goals—in innovative ways that reflect judgements about the skills that are relevant to adult life. (OECD 2001, p. 3)

PISA surveys are conducted every 3 years, with a random sample of 15-year-old students in OECD and partner countries and economies. This age group was chosen because this is around the end of compulsory schooling in many countries. The first PISA survey was in 2000, so that the 2012 survey was the fifth in the series and the sixth is in 2015. This book has been prepared between the data collection for the PISA 2012 survey and the announcement of its first results in December 2013. Further analyses will be published for many years. Every survey administration assesses reading literacy, scientific literacy and mathematical literacy, with a variety of additional assessment components varying across survey administrations such as problem solving, and optional components that also vary such as financial literacy. The meaning of the phrase ‘mathematical literacy’ and the reasons for selecting this as the construct to be assessed in the mathematics component of the PISA survey feature prominently in this book, especially in Chap. 1. In addition to what are usually referred to as the *cognitive* assessment components (the reading, mathematics and science components that relate to recognised and established curriculum domains), background questionnaires directed to schools and students gather data on the school and home environment for learning. Results from PISA are used in many different ways: to compare the performance of students from different countries, to examine the differential performance of students belonging to different subgroups within a country, to track changes in performance over time, and to link features of the learning environment to student performance. Turner and Adams (2007) provide an overview of many organisational and other aspects of PISA.

The surveys are designed so that scores from different survey administrations are directly comparable, so it is now possible to examine trends in achievement over an extended timeframe. In the case of mathematics, the full PISA mathematics scale was developed from the PISA 2003 survey, so mathematics trends can be examined over more than a decade. Because sufficiently many trend items from previous surveys are used within each survey, it is possible to say that a mathematics score of 500 (say) in PISA 2003 describes the same ability level as a score of 500 in PISA 2012. Of course, this is not true for country rankings, because the group of participating countries varies. For example, Finland had a mean score of 536 in PISA 2000 and was ranked fourth. Japan had a mean score of 536 in PISA 2012 and was ranked seventh. The overall performance of Finland in 2000 and Japan in 2012 are the same, with the different rankings reflecting the significant increase in the number of countries participating in PISA over that period.

In each survey administration, the major focus of the survey rotates through reading literacy, mathematical literacy and scientific literacy. The 2003 and 2012 surveys focused on mathematical literacy, with the surveys in 2000, 2006 and 2009 providing a smaller volume of data on mathematics, and with the focus in those years being on either reading or science. For 2003 and 2012, a large number of new mathematical literacy items had to be created and trialled, and this process is described later in this volume in Chap. 6 (by Ross Turner, in discussing test development alongside other aspects of quality assurance in PISA) and Chap. 7 (by Dave Tout and Jim Spithill, from the ACER mathematics test development team for PISA 2012). Mathematics items used in PISA surveys are also presented and discussed in other chapters (including in Chap. 3 by Kaye Stacey as part of her discussion of modelling within PISA mathematics, and in Chap. 8 by Caroline Bardini as part of her discussion of features of the computer-based mathematics option for PISA 2012). In the 2003 and 2012 survey administrations, the questionnaires for students and schools also emphasised mathematics and some of the specifically mathematical probes are discussed in Chap. 10 in this volume by Leland Cogan and William H. Schmidt.

PISA is a huge educational study. In 2012, for example, a random sample of just under 519,000 students in 65 countries (including all 33 of the OECD member countries) participated in the main survey covering mathematics, reading, science, general problem solving and the core background questionnaires, with many undertaking the optional components including computer-based assessment of reading and mathematical literacy, financial literacy, parent questionnaires and student questionnaires on familiarity with ICT and educational careers. All of these instruments were prepared in up to 85 different national versions, including versions translated into 43 different languages using rigorous processes to ensure that they are free from cultural and linguistic biases, so that the data are as truly comparable as possible.

PISA draws on the skills and knowledge of many experts around the world. The 2012 mathematics assessment required 115 new items to be created for the nine-yearly in-depth study of mathematical literacy, alongside 36 items linked to earlier administrations of the survey to enable estimation of trends. From a very large set of raw ideas, new items proposed by teams around the world went into a large pool for extended development and that pool was approximately halved for the field trial and halved again for the main survey in the light of empirical results. Even before selection for the field trial, items were subject to intensive scrutiny by PISA's

Mathematics Expert Group, by the item development teams, by external experts, and by the national teams in every participating country. In 2012, test booklets also included items specially tailored so that emerging economies with currently low performing school systems were able to obtain more reliable data than had been possible in the past.

The substantial length of time between data collection and the release of the first results is in part due to the thorough procedures that are applied to checking the adequacy of the achieved sample of students and schools, and to the sophisticated statistical methods used to produce results, especially in order to make them comparable from survey to survey. To improve the breadth of assessment of mathematical literacy, each student does only a small selection of the full bank of items for mathematics according to a rotated booklet design, within which booklets are assigned randomly to sampled students. Student responses are ‘coded’ according to pre-defined response categories (see Chap. 9 in this volume written by Agnieszka Sułowska).

There is a great deal of information freely available about PISA, past and present, in accordance with OECD policy. The official OECD website (http://www.pisa.oecd.org) includes general descriptions of the project, official reports, links to operational manuals, survey instruments and all released items from previous administrations, and secondary analyses of data on topics of interest. Some other websites, including the website hosted by the Australian Council for Educational Research, which led the international consortium of contractors for PISA from the 2000 to 2012 survey administrations, contain or link to copies of the numerous national and international reports, research publications and commentaries, technical manuals and discussion documents and all released items (e.g. http://pisa-sq.acer.edu.au and http://cbasq.acer.edu.au). It is possible to download databases and manuals for analysis, or to submit a query to an automated data analysis service. In addition to official sites, there are many reports of scientific procedures (e.g. Turner and Adams 2007), secondary analyses of PISA data (e.g. Kotte et al. 2005; Grisay and Monseur 2007; Willms 2010) and many reports with a policy or local focus (see, for example, Oldham 2006; Stacey and Stephens 2008; Stacey 2010, 2011).

A difficult point for mathematics educators to accept is the precise goal of the mathematics work in PISA. All the work carried out to bring PISA mathematics into being is towards the goal of providing the best possible measure of mathematical literacy and its specified components. All of the items are selected on this criterion. Items that do not contribute well are not used, even though they may provide very interesting, important insights into student thinking. Moreover, items have to be coded reasonably economically, so there remains a wealth of information about student performance that is not captured for statistical analysis, although it could possibly be made available for researchers. There are many questions about particular aspects of mathematical thinking where the results of PISA items sometimes provide useful information, but this happens by accident not design, unless it is directly related to the measurement of mathematical literacy.

PISA is not without criticism, but even this can often be seen as a positive result of the OECD’s entry to this space. For example, criticisms have been made of technical and methodological aspects particularly of the analysis of PISA data (for example, see Prais 2003; Goldstein 2004; Kreiner 2011 and a response to Kreiner by Adams 2011). Some criticisms are based on a lack of knowledge of the quality control measures used in item design and survey construction, a gap that this volume hopes to fill. In particular, it is often assumed that no measures are taken to minimise potential biases relating to culture and familiarity with the real-world context. Criticisms have also been made regarding the accessibility of the PISA survey in terms both of the cost of participation and the appropriateness of the test items it uses for countries less wealthy and less developed than most OECD member countries, an important issue that is taken up in several chapters of this volume. A further form of criticism is based on views of the ways in which PISA data are often used, especially where that use is limited to global comparisons of performance with a ‘horse-race mentality’ rather than deeper use to understand the correlates and drivers of performance in order to design system and other educational improvements. It is to be hoped that this volume will play a part in promoting a more informed use of PISA results and constructs. Increased methodological debate related to the conduct of educational surveys might well be seen as a positive outcome of PISA; similarly the number and range of countries either joining PISA or investigating alternative sources of the kind of measures that PISA generates stands as testament to the fundamental importance of the aims of the PISA enterprise. By explaining the inside view of the processes in creating a PISA survey, this book may be seen as a contribution to deepening the nature of consideration and debate about what positive lessons can be learned from PISA and its results.

About This Book

The remainder of this introduction briefly introduces each of the parts of the book in turn. This book is divided into three parts. Part I is concerned with the ideas that are central to PISA mathematics and how they link with other ideas within educational thinking. Part II focuses on the implementation of the survey and Part III brings together perspectives from people around the world who have used PISA initiatives to improve mathematics education in their countries. The chapters differ significantly in style, from broad scholarly surveys and reports of research methods to accounts by individuals of their encounters with PISA ideas and work. Together it is hoped that they provide readers with a rich account of many, but certainly still not all, aspects of the large enterprise that is PISA mathematics.

Part I: The Foundations of PISA Mathematics

Part I reviews the main concepts and theoretical background for the mathematics component of the PISA survey.

Chapter 1 *The Evolution and Key Concepts of the PISA Mathematics Frameworks* by Kaye Stacey and Ross Turner describes the key concepts of the Frameworks and some of the history and origins of those ideas, within PISA and from broader educational thinking.

Chapter 2 *Mathematical Competencies and PISA* by Mogens Niss describes the origins of a set of mathematical competencies that take a central place in the PISA Framework to describe what it means to 'do mathematics'.

Chapter 3 *The Real World and the Mathematical World* by Kaye Stacey describes how PISA theorises the link between mathematics and its use for practical purposes through the mathematical modelling cycle, and how an assessment using real-world contexts can be implemented fairly across groups and cultures.

Chapter 4 *Using Competencies to Explain Mathematical Item Demand*: *A Work in Progress* by Ross Turner, Werner Blum and Mogens Niss describes research that has shown how the PISA mathematical competencies can be used to understand aspects of the cognitive demand of PISA mathematics tasks.

Chapter 5 *A Research Mathematician's View on Mathematical Literacy* by Zbigniew Marciniak presents a personal reflection on how involvement with PISA mathematics has affected his views about what is important in mathematics education. Including this reflection acts as a reminder that important theoretical considerations actually have an impact on individuals involved in education. The issues that it addresses have been at the heart of the 'math wars' that have raged in many countries over several decades.

Part II: Implementing the PISA Survey: Collaboration, Quality and Complexity

Part II describes aspects of the implementation of the PISA survey from various insider perspectives, showing the complexity of the PISA enterprise, the steps taken to ensure quality of PISA outcomes and the extensive collaboration among a variety of stakeholders and other players that takes place to make the enterprise such a success.

Chapter 6 *From Framework to Survey Data*: *Inside the PISA Assessment Process* by Ross Turner introduces the major elements involved in the development and implementation of each PISA survey.

Chapter 7 *The Challenges and Complexities of Writing Items to Test Mathematical Literacy* by Dave Tout and Jim Spithill provides an outline of the processes of test development. It uses released PISA items to exemplify the processes.

Chapter 8 *Computer-Based Assessment of Mathematics in PISA 2012* by Caroline Bardini describes theoretical and practical issues related to the computer delivery of PISA items and illustrates with several of the PISA 2012 items.

Chapter 9 *Coding of Mathematics Items in the PISA Assessment* by Agnieszka Sułowska provides a very practical account of the way student responses to PISA items are processed from the perspective of a PISA national assessment centre.

Chapter 10 *The Concept of Opportunity to Learn* (*OTL*) *in International Comparisons of Education* by Leland Cogan and William Schmidt discusses the development and inclusion of innovative questions related to opportunity to learn mathematics in the student questionnaire for PISA 2012.

After the stages of item creation, the data collection and the coding that are described in these chapters, a long and complex process of collating, cleaning, processing and then reporting results ensues. Understanding the statistical procedures is also important to a well-informed interpretation of the PISA results. This is beyond the scope of this book, but is well described in the technical manuals written as part of the documentation for each PISA survey administration (e.g. Adams and Wu 2002).

Part III: PISA's Impact Around the World: Inspiration and Adaptation

Part III of the book is a collection of reflections on the impact that PISA has had on individuals' thinking, on education systems, and on teaching and learning practice in 14 different countries.

Chapter 11 *Applying PISA Ideas to Classroom Teaching of Mathematical Modelling* by Toshikazu Ikeda discusses the application of the ideas related to mathematical modelling, as promoted in the PISA Framework, in classroom practice in Japan.

Chapter 12 *The Impact of PISA on Mathematics Teaching and Learning in Germany* by Manfred Prenzel, Werner Blum and Eckhard Klieme, discusses the changes instituted in German schools and systems as a direct consequence of concern about Germany's unexpectedly low initial PISA results.

Chapter 13 *The Impact of PISA Studies on the Italian National Assessment System* by Ferdinando Arzarello, Rossella Garuti and Roberto Ricci describes efforts to reform classroom practices in order to better prepare Italian students for the kinds of thinking valued through PISA.

Chapter 14 *The Effects of PISA in Taiwan*: *Contemporary Assessment Reform* by Kai-Lin Yang and Fou-Lai Lin describes contested plans in a high-performing PISA country to introduce reforms arising from Taiwan's PISA results.

Chapter 15 *PISA's Influence on Thought and Action in Mathematics Education*, compiled by Kaye Stacey, is a collection of shorter pieces that provide reflections on aspects of the impact of PISA in ten countries. It speaks to the influence of PISA ideas around the world as well as to its congruence with the major concerns of many educators.

Final Reflections

Compiling this book marks the end of a long process for those of us working on mathematics for PISA 2012. The framework has been revised so that it better shows the connections among its elements, with work by the Mathematics

Expert Group (MEG), ACER, and *Achieve*, and with input from experts around the world. Organised by ACER, a huge number of items were developed by teams around the world, critiqued numerous times including by teams in all countries, selected by the MEG, translated into 43 languages, administered and coded in the field trial in 65 countries and statistically analysed to provide data for selection of items into the main survey, involving its major administration, coding, statistical analysis and finally presentation of the first results in December 2013. For most people, PISA begins at this point when the first results are available.

These results are only worth the investment of so much effort if they can be used for productive purposes to improve educational outcomes. This in turn depends on the extent to which the processes that are followed provide confidence in the reliability and integrity of the results, and whether PISA outcomes generate insight into student performance. Whereas dealing with complexity is one theme of many of the chapters, using strong quality assurance measures is another. We hope that some of the qualms about PISA's capacity to provide good measures across countries will be alleviated by reading this book.

As is evident in many chapters of this book, the concept of mathematical literacy is well founded within the tradition of mathematics education but is also a distinct new contribution, especially because the PISA processes have forced some integration of analyses from across the globe. Around the world, countries are adopting, and of course adapting, mathematical literacy as the major goal of schooling for most students. Unlike some of its variants such as numeracy and despite the impression given in some countries as a result of the words typically used to render the terminology in different languages, mathematical literacy for all is not a low level goal, but a high aspiration. In Part III, there are examples of countries where PISA's analysis of mathematical literacy is also forming a framework for national curriculum development. The pool of publicly released items from PISA Mathematics is now sizeable, and is also beginning to be used quite widely, in educational research, for teacher professional development and as a model for assessment. In other words, PISA has grown out of existing traditions and practices in mathematics education, and in turn has influenced the directions in which mathematics education is developing. Whilst close copying of PISA-style items is not sufficient to encourage strong mathematical literacy, because the inevitable constraints of providing robust international assessment limit the range of such tasks, the released items certainly provide ideas and directions for improving instruction. They can also stimulate the production and classroom use of PISA-like tasks that develop mathematical modelling and mathematical literacy more richly. PISA also publishes all databases, so further research on many fronts is supported. With well-informed commentators, PISA can make a contribution far beyond the horse race results so frequently represented in the media.

Working within PISA also makes it clear that one of the world's largest educational research surveys cannot answer all of the research questions that need to be answered. It cannot, for example, directly answer questions about the best direction for educational reform. However, we hope that this book contributes to widespread better understanding of PISA results, so that they can be sensibly used as a basis for the needed experimentation, study and policy development that can follow having the strong measure of mathematical literacy that PISA surveys provide.

Melbourne Graduate School of Education
The University of Melbourne
Melbourne, VIC, Australia
Kaye Stacey

International Surveys, Educational
Monitoring and Research
Australian Council for Educational Research
Melbourne, VIC, Australia
Ross Turner

References

Adams, R. (2011). Comments on Kreiner 2011: Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment. http://www.oecd.org/pisa/47681954.pdf. Accessed 3 Dec 2013.

Adams, R., & Wu, M. (Eds.). (2002). *The PISA 2000 technical report*. Paris: OECD Publications.

Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education*: *Principles*, *Policy & Practice, 11*(3), 319–330. doi: 10.1080/0969594042000304618

Grisay, A., & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, *33*, 69–86.

Kreiner, S. (2011). *Is the Foundation Under PISA Solid*? *A critical look at the scaling model underlying international comparisons of student attainment*. University of Copenhagen: Department of Biostatistics.

Kotte, D., Lietz, P., & Lopez, M. (2005). Factors influencing reading achievement in Germany and Spain: Evidence from PISA 2000. *International Education Journal*, *5*(4), 113–124.

Organisation for Economic Co-operation and Development (OECD). (2001). *Knowledge and skills for life*: *First results from PISA 2000*. Paris: OECD Publications.

Oldham, E. (2006). The PISA mathematics results in context. *The Irish Journal of Education* (*Iris Eireannach an Oideachais*), *37*, 27–52.

Prais S. J. (2003). Cautions on OECD's recent Educational Survey (PISA). *Oxford Review of Education*, *29*, 139–163.

Stacey, K. (2010). Mathematical and scientific literacy around the world. *Journal of Science and Mathematics Education in Southeast Asia*, *33*(1), 1–16. http://www.recsam.edu.my/R&D_Journals/YEAR2010/june2010vol1/stacey(1-16).pdf.

Stacey, K. (2011). The PISA view of mathematical literacy in Indonesia. *Journal of Indonesian Mathematics Society B* (*Journal on Mathematics Education*), *2*(2), 95–126.

Stacey, K., & Stephens, M. (2008). Performance of Australian school students in international studies in mathematics. *Schooling Issues Digest* 2008/1. Canberra: Department of Education, Employment and Workplace Relations.

Turner, R., & Adams, R. (2007). The programme for international assessment: An overview. *Journal of Applied Measurement*, *8*(3), 237–248.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record*, *112*(4), 1008–1037.

Part I
The Foundations of PISA Mathematics

Introduction to Part I

In this part the inside story of the conceptualisation of mathematical literacy for PISA for the first five surveys is presented. Authors have been directly involved in creating the PISA Mathematics Framework, which specifies the assessment parameters and the nature of the mathematical literacy items. The key elements of the Mathematics Framework for PISA 2012 are introduced in the context of a discussion of the evolution of the Frameworks of the PISA survey from 2000. The relationships between the *literacy* notion and other ideas underpinning the PISA Framework, and the appearance of similar ideas elsewhere in the mathematics education world show clearly that the developments here form part of an ongoing historical progression in the thinking of policy-makers and educational practitioners of all kinds that is aimed at improving the quality of mathematics education.

In Chap. 1, Kaye Stacey and Ross Turner put the 2012 Framework in its historical context, emphasising the links between ideas harnessed in this Framework and other contexts in which the same or similar ideas have been used. Two major sets of ideas central to PISA mathematics since its inception are mathematical modelling and mathematical competencies.

In Chap. 2, Mogens Niss provides an extensive history of the development of the competency notion for mathematics, which is the general attempt to describe mathematics in terms of a small set of competencies involved in doing mathematics rather than by naming the topics studied in mathematics courses. The importance of this is that it focusses the attention of teachers, assessors and students on working mathematically in the broadest sense, not just on knowing how to solve routine problems. This description of competencies centres on the work of Niss himself and colleagues in Denmark but links to some other well-known schemes are also discussed. The chapter reports on the evolution of the competencies (renamed the fundamental mathematical capabilities for PISA 2012) over the first fifteen or so years of PISA's existence from the more general schemes to one specifically designed for PISA purposes. In conjunction with Chap. 4, this chapter offers an accessible and authoritative outline of the history, background and current developments of these influential ideas.

In Chap. 3, Kaye Stacey explains how mathematical modelling and mathematisation fit within PISA mathematics, using a number of released PISA items to illustrate the points made. The central idea of mathematical literacy is that is it about the use of mathematics in people's lives, and this raises issues of authenticity and interest of the real-world contexts and the equity of assessment using them. Assessing mathematics in context is more complicated than assessing mathematical skills and routines. A further contribution of the chapter is in clarifying the meaning and use of many different terms (such as literacy, numeracy, competency, modelling, mathematising) that are sometimes used in discussions about PISA.

In Chap. 4, Turner, Blum and Niss present the story of ongoing research that has exposed aspects of the role played by mathematical competencies in affecting the empirical difficulty of PISA items, and therefore the expression of the literacy construct of which PISA items are intended to provide indicators. The chapter elaborates on the definition and operationalisation of the competencies and how this has been used in task development. The detailed discussion of the thinking behind the scheme and its modifications is invaluable for anyone aiming to understand the role of competencies in doing mathematics. The final appendix, which defines the competencies and the specifications of four levels for each, is a definitive guide for researchers, teachers and test designers intending to use competencies to explain, monitor or manipulate item demand.

Marciniak completes this part by providing in Chap. 5 a personal reflection from the perspective of a pure mathematician on the changes in his thinking about mathematics education that have resulted from his grappling with the main ideas and practices of PISA mathematics, first as a national reviewer of draft PISA material, and then as a member of the Mathematics Expert Group for the last four survey administrations. The contribution belongs in this part because it is intimately about what PISA should value most. Marciniak reflects on his growing realisation that for most students at school, the goal of mathematical literacy is of greater importance than promoting abstract mathematical thinking, and that the common 'catch the fox' approach to curriculum does not serve students well. This is an individual account, contrasting in style to other chapters in this part, but it is significant because of ongoing community debate about what should be the highest priorities of school mathematics, and hence what type of mathematics PISA should assess. Whilst the 'math wars' (Schoenfeld 2004) in the USA are extensively documented aspects of this debate, many educators and professional mathematicians around the world grapple with this issue. The beauty and structure of pure mathematics and the opportunities for truly challenging problem solving attracted many of us (including Marciniak) to work in mathematics, but mathematics as a compulsory subject must place the highest priority on its usefulness.

There is a humorous saying in English that 'a camel is a horse designed by a committee'. As readers of this part encounter some of the extra camel humps in the conceptual framework of PISA, they will see the signs that PISA has been designed by numerous committees, modified over time, and has taken on board suggestions from around the world. But just as a real camel has characteristics that make it a valuable and unique animal and not just a poorly designed horse, the PISA 'camel' is a strong and robust beast, fit to withstand the many perils in the desert of international assessment. It has been designed through genuine collaborative thinking, rather than bureaucratic committee processes. It has amalgamated constructs and ideas from many sources, expressed in many different educational traditions and languages, to build a framework upon which an assessment of valuable learning for citizens around the world can be founded.

Reference

Schoenfeld, A. H. (2004). The math wars. *Educational Policy, 18*(1), 253–286. doi:10.1177/0895904803260042.

Chapter 1
The Evolution and Key Concepts of the PISA Mathematics Frameworks

Kaye Stacey and Ross Turner

Abstract This chapter describes the purpose of the Framework for the PISA surveys of mathematical literacy and its evolution from 2000 to 2012. It also describes some of the analysis and scholarship on which the key constructs of the Framework are based, and links to kindred concepts in the wider mathematics education literature. The chapter does not intend to present the Framework but instead to share insights into its creation by successive Mathematics Expert Groups. The main Framework concept is that of mathematical literacy which has its roots in recognition of the increasing importance of mathematical proficiency in the modern world. The chapter describes mathematical literacy, its evolving definition and the origin of the term within broadened notions of literacies and its relationship to other terms such as quantitative literacy and numeracy. It describes the central constructs of the Framework, which are used to describe what abilities make up mathematical literacy and are also used to ensure that the item pool is comprehensive and balanced. These are the real-world context categories that group the source of mathematical challenges, the phenomenologically-based content categories, the fundamental mathematical capabilities and a set of processes based on the mathematical modelling cycle. The way in which new technologies have expanded the view of mathematical literacy and how this has been assessed through the 2012 computer-based assessment of mathematics is also discussed.

K. Stacey (✉)
Melbourne Graduate School of Education, The University of Melbourne,
234 Queensberry Street, Melbourne, VIC 3010, Australia
e-mail: k.stacey@unimelb.edu.au

R. Turner
International Surveys, Educational Monitoring and Research, Australian Council
for Educational Research (ACER), 19 Prospect Hill Rd, Camberwell, VIC 3124, Australia
e-mail: Ross.Turner@acer.edu.au

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_1

Introduction

Imagine you were asked to find out whether educational systems around the world are doing a good job in preparing students for the challenges that they are likely to face in their futures. You are almost certain to decide that the traditional ‘three Rs’—reading, ’riting and ’rithmetic—remain highly important, along with other capabilities about which there will be more debate. Now focus on arithmetic. Here, you are likely to decide that restricting your investigation to arithmetic is definitely out of date and that you need to investigate success in the broad field of mathematics. (Here, and almost everywhere else in this volume, this term ‘mathematics’ includes all branches of the mathematical sciences, including statistics.) This needed breadth has been recognised for many years. For example, in 1989 the National Council of Teachers of Mathematics commented:

> To become mathematically literate, students must know more than arithmetic. They must possess a knowledge of such important branches of mathematics as measurement, geometry, statistics, probability, and algebra. These increasingly important and useful branches of mathematics have significant and growing applications in many disciplines and occupations. (NCTM 1989, p. 18)

Within this wide domain of mathematics, what sort of tasks should be posed to answer the main concern of the OECD’s PISA survey for mathematics: have students have been well prepared mathematically for future challenges (OECD 2000)? The main topic of this chapter is to discuss the PISA answer to this question: that the highest priority for assessment is ‘mathematical literacy’ with its focus on life after school, not just life at school. The chapter discusses the concept of mathematical literacy from many points of view, including its history from before PISA and as it developed through the 2000–2012 surveys. It provides an analysis of the components of mathematical literacy (and their origins in many branches of educational thought) and describes how this analysis is employed to create a balanced assessment of mathematical literacy. The way in which PISA operationalises these components of mathematical literacy is officially described in the Mathematics Framework (see, for example, OECD 2013a), so the chapter begins with a brief description of its purpose and history.

The Frameworks from PISA 2000 to PISA 2012

Much of the subsequent discussion in this chapter draws on the Frameworks for mathematics for the PISA surveys from 2000 to 2012 (OECD 1999, 2004, 2006, 2009c, 2013a). These were created by the Mathematics Expert Groups (MEG) appointed for each survey by the international contractors with the approval of the PISA Governing Board. MEG members include mathematics educators, mathematicians and experts in assessment, technology, and education research from a range of countries. The preface lists the membership from 2000 to 2012. External review of the Frameworks has been widely sourced over time, with the U.S.A. group *Achieve* (www.achieve.org) co-ordinating major input for the PISA 2012 Framework. Whilst the Mathematics Framework has been revised and published anew for each administration of the PISA survey, only the initial Framework (OECD 1999) and the versions for PISA 2003 and PISA 2012 (OECD 2004, 2013a) when mathematics was the major survey domain represent significant developments.

The purpose of the Framework is to set out the PISA approach and describe the assessment instruments in terms of the processes that students need to perform, the mathematical content that is relevant, and the real-world contexts in which knowledge and skills are applied. This analysis of the concept of mathematical literacy and what contributes to student success is used to ensure that the assessment gives a sufficiently balanced and thorough coverage of the domain to gain the support of countries participating in the PISA survey. The Mathematics Framework also identifies mathematics-related aspects of the assessment of attitudes that contribute to students using and further developing their capabilities.

The Frameworks for the first four surveys were developed by the MEGs under the chairmanship of Professor Jan de Lange from the Netherlands. de Lange's leadership provided a strong link to the Freudenthal Institute's approach to mathematics education, known widely as Realistic Mathematics Education (RME). The first Framework was only partially developed, but it made a clear statement of the centrality in PISA of the mathematisation of the real world that permeates de Lange's RME perspective (de Lange 1987). PISA was therefore able to capitalise on an existing body of research and resources (see, for example, de Lange 1992). A more complete development was undertaken for PISA 2003 and this second Framework (OECD 2004) began to flesh out the description of the process of doing mathematics and the competencies involved. The changes that were made to the Frameworks for the 2006 and 2009 survey administrations were largely cosmetic, but when mathematics was again the major survey domain for PISA 2012, the Framework (OECD 2013a) underwent a major revision. This chapter is intended as a behind-the-scenes explanation of framework ideas: the published Frameworks remain the authoritative source of the outcomes of that development.

What Is Mathematical Literacy?

The task for PISA, as set by the OECD is to discover whether students have been well prepared mathematically for future challenges in life and work. What sort of mathematical tasks should be posed to answer this question? Consider Pythagoras's theorem, arguably the most important theorem in all of mathematics, known for over 3,000 years. It provides practical information for calculating distances and it is used and generalised in many different branches of pure and applied mathematics. It has about 370 known proofs. It also motivated Fermat's Last Theorem, the most famous of all mathematical problems. Certainly knowledge of Pythagoras's

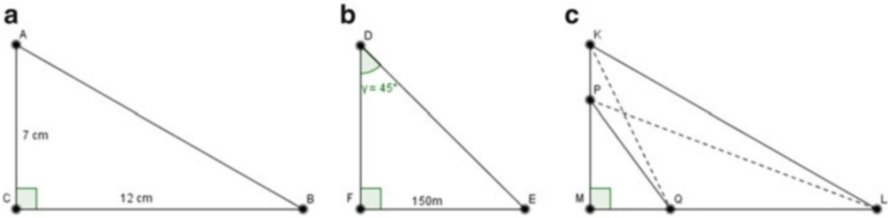


Fig. 1.1 Diagrams for sample problems involving Pythagoras's theorem

theorem is important, but what type of problems about it would be appropriate to ask? Figure 1.1 offers a range of possibilities. All of these are valid questions that students could be asked at school when studying Pythagoras's theorem.

- Sample Problem 1. State Pythagoras's theorem.
- Sample Problem 2. ABC (see Fig. 1.1a) is a triangle right-angled at C. AC has length 7 cm. BC has length 12 cm. Calculate the length of side AB.
- Sample Problem 3. In triangle DEF (see Fig. 1.1b), angle F is 90°, angle D is 45° and side EF is 150 m. Calculate the length of side DE.
- Sample Problem 4. A large kite is flying at an angle of 45° to the ground at height of 150 m. How long is the rope tethering it?
- Sample Problem 5. KLM (see Fig. 1.1c) is a triangle right-angled at M. P is a point on KM and Q is a point on LM. Prove that $KQ^2 + LP^2 = KL^2 + PQ^2$.
- Sample Problem 6. Prove Pythagoras's theorem.

Sample Problem 1 tests recall of fundamental knowledge that is required to answer all of the sample problems that follow. Sample Problem 2 is a very straightforward application of the theorem also requiring accurate calculation. Sample Problem 3 draws in other geometric knowledge (triangle DEF is isosceles, and so has two equal sides) before the knowledge of Pythagoras's theorem as tested in Sample Problem 2 can be used. Sample Problem 4 has the same mathematical core as Sample Problem 3, but is presented in a context. Thus the problem solver first has to uncover the mathematical structure within the real-world situation described, introducing for himself or herself the triangle and the right angle using real-world knowledge and deciding whether it is reasonable to consider the rope as a straight line, at least as a first approximation. As with Sample Problem 3, the intra-mathematical Sample Problem 5 requires devising a problem solving strategy, although in this case it does not draw in knowledge beyond Pythagoras's theorem. Instead it requires the insight that Pythagoras's theorem can be used in four different right-angled triangles within the figure, followed by use of a little algebra.

Like Sample Problem 5, Sample Problem 6 is again in the intra-mathematical world, connecting students' experience to the great advance that the Pythagoreans are credited with. They changed mathematics from the practice of rules for numerical calculation to an intellectual structure by "examining its principles from the beginning and probing the theorems in an immaterial and intellectual manner" (Boyer 1968, p. 53). Depending on students' mathematical experience, Sample

Problem 6 may be answered by reproduction of 'book knowledge', or it may present a substantial challenge. Questions like all of those above could potentially be asked to investigate the effectiveness of educational systems.

In preparation for the first PISA assessment, the OECD and its Framework developers needed to decide what subset and style of mathematics was the most important for PISA to assess. The answer was summarised in the phrase 'mathematical literacy'. The key idea is to assess as directly as possible students' ability to use mathematics in solving problems arising in authentic real-world problems, rather than to make unsupported inferences about that ability by examining only the abstracted core mathematical knowledge and skills. The PISA 2000 report explains that the term 'literacy' is used

> to indicate the ability to put mathematical knowledge and skill to functional use rather than just to master it within a school curriculum. (OECD 2000, p. 50)

Sample Problem 4 above is the closest to a PISA problem; in fact it is an abbreviated version of an item from the PISA 2012 main survey, PM923Q03 Sailing ships Question 3, shown in Fig. 1.2. The Skysails Company http://www.skysails.info/english/power/ makes sails to supply green power from the wind to drive ships and for power generation at sea. This authentic situation provides the stimulus for items involving percentage change (PM923Q01 Question 1), real-world interpretation of algebraic formulas (PM923Q02 Question 2 not released), Pythagoras's theorem (PM923Q03 Question 3) and a multi-step calculation involving rates (PM923Q04 Question 4). Like Sample Problem 4, solving PM923Q03 Sailing ships Question 3 involves creating a mathematical model of the real situation and then applying the same intra-mathematical thinking as in Sample Problem 3 above, which in turn involves the component knowledge and skills of Sample Problems 2 and 1. Items that test mathematical literacy involve the creation, use or interpretation of a mathematical model for a real-world problem as well as intra-mathematical thinking. PISA does not set out to test 'book knowledge' or factual recall, except as part of solving a problem in an authentic situation, although in some of the simplest items the real situation is, in fact, involved in only a minimal way. These ideas are discussed fully in Chap. 3 of this volume.

A Continuum to Complex Mathematical Thinking

Can questions testing mathematical literacy involve intra-mathematical thinking and proof of the complexity of Sample Problem 5 or Sample Problem 6 above? Producing insightful solutions to complex problems can be part of mathematical literacy, provided the need for the thinking emerges from a realistic context and solving the problem could genuinely describe, explain or predict something about that context. Mathematical literacy can also involve the presentation of convincing arguments about those real situations, and the special proof-related nature of these is a characteristic of mathematics.

SAILING SHIPS

Ninety-five percent of world trade is moved by sea, by roughly 50 000 tankers, bulk carriers and container ships. Most of these ships use diesel fuel.

Engineers are planning to develop wind power support for ships. Their proposal is to attach kite sails to ships and use the wind's power to help reduce diesel consumption and the fuel's impact on the environment. (original image not shown) (layout changed from original)

Question 1: SAILING SHIPS *PM923Q01*

One advantage of using a kite sail is that it flies it flies at a height of 150 m. There, the wind speed is approximately 25% higher than down on the deck of the ship.

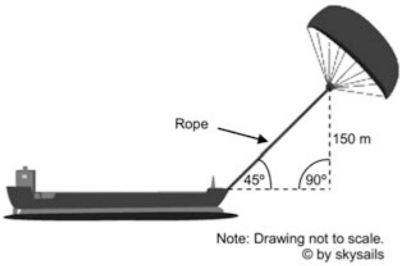
At what approximate speed does the wind blow into a kite sail when a wind speed of 24 km/h is measured on the deck of the ship?

A. 6 km/h
B. 18 km/h
C. 25 km/h
D. 30 km/h
E. 49 km/h

Question 3: SAILING SHIPS *PM923Q03*

Approximately what is the length of the rope for the kite sail, in order to pull the ship at an angle of 45° and be at a vertical height of 150 m, as shown in the diagram opposite?

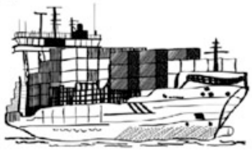
A. 173 m
B. 212 m
C. 285 m
D. 300 m



Question 4: SAILING SHIPS *PM923Q04*

Due to high diesel fuel costs of 0.42 zeds per litre, the owners of the ship *NewWave* are thinking about equipping their ship with a kite sail.

It is estimated that a kite sail like this has the potential to reduce the diesel consumption by about 20% overall.



Name: *NewWave*	Load capacity: 12 000 tons
Type: freighter	Maximum speed: 19 knots
Length: 117 metres	Diesel consumption per year without a kite sail:
Breadth: 18 metres	approximately 3 500 000 litres

The cost of equipping the *NewWave* with a kite sail is 2 500 000 zeds.

(layout condensed)

After about how many years would the diesel fuel savings cover the cost of the kite sail? Give calculations to support your answer.

Fig. 1.2 PM923 Sailing ships, released after PISA 2012 main survey (OECD 2013b)

When discussing complex mathematical thinking, an important caveat for the implementation of PISA is that the questions are able to be solved by an adequately large percentage of the target age group, under the conditions in which the survey is administered. It is useless to include in the PISA survey questions with very high or very low success rates because an item makes very little contribution to the measurement if nearly all students obtain the same score. As a construct, there is no bound to the complexity of mathematical literacy items and it transcends age boundaries, but the items used in the PISA survey must take the characteristics of 15-year-old students into account. Only a subset of mathematical literacy items can be used with 15-year-olds.

Mathematical literacy, as defined by PISA, is not something that people have or do not have, instead it is something that everyone possesses to a greater or lesser degree. Proficiency lies along a continuum applying to very direct, simple tasks in everyday situations through to situations involving the highest levels of technical work. As noted by Marciniak (Chap. 5 of this volume), when judging the appropriateness of the mathematical content for PISA items, it is more important to select items involving content that features prominently in functional use than advanced, difficult content.

Formal Definitions

For PISA 2000 mathematical literacy was defined as:

> an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to engage in mathematics, in ways that meet the needs of that individual's current and future life as a constructive, concerned and reflective citizen. (OECD 1999, p. 41)

For PISA 2006 mathematical literacy was revised to:

> an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen. (OECD 2006, p. 72)

The definition has again been revised for the 2012 Framework (OECD 2013a) but in all of these revisions, there has not been an intention to change the underlying construct. For 2012 the revision, in response to international comment, was intended to clarify the ideas underpinning mathematical literacy so that they can be more transparently operationalised and to identify more clearly the fundamental and growing role that mathematics plays in modern society. The formal PISA 2012 definition of mathematical literacy is as follows:

> 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. (OECD 2013a, p. 25)

All of these definitions are built on the consensus of the governments supporting PISA and most research literature that all adults, not just those with technical or scientific careers, now require a more sophisticated level of mathematical literacy than in the past (see, for example, Autor et al. 2003).

The first sentence of the 2012 definition identifies mathematical literacy as a capacity of individuals and asserts the centrality of working in context, as described above. It asserts that mathematical literacy is very closely related to mathematical modelling, because formulating mathematical models, employing mathematical knowledge and skills to work on the model and interpreting and evaluating the outcome are its essential processes. The second sentence explains that all aspects of mathematics are involved in mathematical literacy, whether through specific mathematical concepts and techniques or generic mathematical reasoning. The definition also highlights the functional purpose of mathematical literacy: to increase understanding of real-world phenomena and hence to support sound decision making across all areas of life. This is not a new idea. One of the reports that followed the release of the PISA 2003 outcomes (OECD 2009b) cites Josiah Quincy writing in 1816 of the importance of 'political arithmetick' to fulfil the duties of a citizen conscientiously. Both the published Framework for PISA 2012 (OECD 2013a) and Stacey (in press) unpack further aspects of this definition.

Why Call It 'Mathematical Literacy'?

The name 'mathematical literacy' has come to be associated with PISA, as part of its broadened understanding of literacy in modern society (OECD 1999), but it has a longer history. Ray Adams, the International Project Director contracted by the OECD to lead development and implementation of the first five PISA survey administrations, reminisced that he suggested the name 'mathematical literacy' at the beginning of work on PISA (as part of the broad notion of literacy as described below for all PISA domains) but he does not recall a specific source. In fact, the phrase was already being used, although it was not widespread. Turner (2012) points to usage in the 1940s without definition. The introduction to the famous NCTM Standards (National Council of Teachers of Mathematics 1989) reports how they began with a Commission charged with creating "a coherent vision of what it means to be mathematically literate" (p. 1) and went on to summarise the term as denoting:

> an individual's ability to explore, to conjecture, and to reason logically, as well as to use a variety of mathematical methods effectively to solve problems. By becoming literate, their mathematical power should develop. (NCTM 1989, p. 6)

This early definition includes two features of the use of the word 'literacy': that it involves functional use of knowledge (applying knowledge to solve problems—by implication important problems) and that it increases the individual's power. Comber (2013), writing on the development of the reading-writing concept of literacy and its relation to critical theory, notes that the term 'literacy' did not come into common use until the middle of the twentieth century and then it was used, especially as illiteracy, mainly in adult education and to describe the needs of the developing world. She reports how critical theorists such as Paulo Friere changed the view of literacy away from a skill to be mastered and instead put the emphasis on engagement with the world in the expectation that literacy should transform workers' lives. When we translate into the context of mathematics, this means to have the power to use mathematical thinking to solve real-world problems to better deal with the challenges of life. The 2003 PISA Mathematics Framework (OECD 2004) takes up the distinction made by Gee (1998) between the design features of a language (e.g. its grammar) and the social functions of language. It makes a parallel distinction between design features of mathematics (concepts, procedures, conventions) and the functions that mathematics can serve in the wider world. Like Gee, PISA emphasises how education must not focus on the design features to the exclusion of the function. This is a broad theme across mathematics education although rarely expressed in those terms.

Adopting the term 'mathematical literacy' was also strongly influenced by the long-standing use of the term 'scientific literacy'. Bybee (1997) provides a brief history, dating 'scientific literacy' back to at least the 1950s. It denotes a familiarity with science on the part of the general public and an orientation to helping people understand the world they live in and to act appropriately (DeBoer 2000). It is part of a push for a broad school treatment of science and its implications for society. Turner (2012) gives a broad discussion of the links to scientific literacy, as well as to the concepts that are discussed in the next section.

By 2012, mathematical literacy has become a common phrase: a search of the index of the electronic pre-proceedings of the 2012 International Congress for Mathematical Education showed it was used in 10 % of the 500 submitted papers.

Mathematical Literacy, Numeracy and Quantitative Literacy

There are at least two other terms in widespread use with strong links to mathematical literacy: numeracy and quantitative literacy. Neither of these has a universally agreed definition. One advantage of PISA's use of the initially less familiar term 'mathematical literacy' is that consistent use of the PISA definition might contribute to better communication within mathematics education.

The term 'numeracy' has been principally used in countries influenced by the United Kingdom where it was coined as a mirror image to literacy in the Crowther Report of 1959 with a broad meaning (Cockcroft 1982), quite closely related to mathematical literacy. The influential Cockcroft Report noted a narrowing of the term by 1982, and described the goal of numeracy as "an 'at-homeness' with numbers and an ability to cope confidently with the mathematical demands of everyday life" along with "an appreciation and understanding of information which is presented in mathematical terms" (Cockcroft 1982, para 39, p. 11). It went on to give a list of mathematics topics for lower achieving students, to be taught alongside a range of applications. Numeracy continues to be used in several different senses: as a minimum expectation for the mathematical knowledge of all learners so that they can cope in the world, as a label for the mathematics learned in the early years of school (especially in the Number domain), or as a solid foundation for meeting the mathematical demands of higher education and most work. In summary, some uses of the term 'numeracy' are very close to PISA's 'mathematical literacy' and others are far away.

The report of the first OECD Adult Literacy Survey (OECD 1995) explains that it follows earlier practice in dividing literacy into three domains: prose literacy, document literacy and quantitative literacy. Quantitative literacy is described as:

> the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a cheque book, figuring out a tip, completing an order form or determining the amount of interest on a loan from an advertisement. (OECD 1995, p. x)

This closely defined interpretation of 'quantitative literacy' contrasts with broader uses of the term, especially in the U.S.A. such as that of the influential report "Mathematics and Democracy: The Case for Quantitative Literacy" (Steen 2001). This describes examples across a wide range of aspects of life (e.g. citizenship, personal finance, education, management) and skills drawing on understanding of broadly interpreted branches of mathematics (e.g. arithmetic, data, computers, statistics, modelling). It is close to PISA's mathematical literacy. The essential role of context in quantitative literacy is reiterated in many places in the book, as in this passage:

> . . . mathematics focuses on climbing the ladder of abstraction, while quantitative literacy clings to context. Mathematics asks students to rise above context, while quantitative literacy asks students to stay in context. Mathematics is about general principles that can be applied in a range of contexts; quantitative literacy is about seeing every context through a quantitative lens. (Hughes-Hallett 2001, p. 94)

Confusingly, the Steen report seems to use the terms 'quantitative literacy' and 'numeracy' synonymously. It sometimes uses the term 'mathematical literacy' to relate only to intra-mathematical tools and vocabulary but elsewhere conveys the PISA meaning. However, the report also contains a useful discussion of the origins of all the terms, as does Turner (2012). In yet another variation, de Lange (2006) sees the relationship somewhat differently with mathematical literacy the overarching concept, having subsets of quantitative literacy, spatial literacy and numeracy, and the PISA phenomenological content categories contributing in different ways to each these literacies.

The major difficulty with all of these words is that sometimes people use them in a narrow sense, so that the broad ambitious sense of PISA's mathematical literacy, for example, is often not appreciated. This is an especially serious issue in some languages. A strong criticism of the name 'mathematical literacy' comes from countries particularly in the Spanish speaking world, but in other places too, where the word 'literacy' has such an entrenched narrow meaning in their language that it can be impossible to convey the broader meaning intended by PISA in local and national educational debates. As Professor Maria Sánchez has put it in a personal communication to Ross Turner (and cited in Turner 2012)

> The word for 'literacy' in Spanish is 'alfabetización'. This concept leads to very basic reading and writing abilities. So 'alfabetización matemática' would be interpreted as knowing how to count and add, more or less, but no more than that.

The response in Uruguay, for example, to PISA's use of the name 'mathematical literacy' was to refer initially to 'mathematical culture', 'scientific culture' and 'reading comprehension'. More recently the concepts of 'cognitive competency', 'cognitive processes', 'developing of competencies for life', have gained wider acceptance, so they now refer to *Competency in Mathematics, in Science and in Reading*. The French language has a similar difficulty with the term literacy because the translation to 'alphabétisation' is narrow and very strongly linked to reading and writing. Instead the term 'culture mathématique' is now being used in reports from the French government such as that by Keskpaik and Salles (2013). Keskpaik and Salles define 'la culture mathématique' by translating the official PISA 2012 definition for mathematical literacy given above.

The international concerns have led to pressure to modify the PISA language. So at the organisational level, the OECD has shifted its language towards referring to PISA as an assessment of mathematics, science and reading; and where reference to 'mathematics' is not sufficient, to refer to 'mathematical competence'. This is intended to convey the same meaning as *mathematical literacy* but aims to avoid the narrow connotations of that term. Nevertheless, within each of the survey domains, the *literacy* reference has been retained at least in English and in languages that do not have such a strong association of literacy with only a basic level of understanding. There is a possibility that the formal name may change in the future: the Context Questionnaire Framework for PISA 2012, for example, uses the phrase 'mathematical competence' instead (OECD 2013a, p. 183).

Mathematics and Mathematical Literacy: Set or Subset?

The quote from Hughes-Hallett above raises the question of whether mathematical literacy, along with quantitative literacy and numeracy, are best considered as a part of mathematics, or whether they are best considered as being larger than mathematics or just different to it. For those who think that 'mathematics' is best constrained to the abstract and theoretical, mathematical literacy interpreted broadly must go beyond mathematics, because mathematical literacy tasks involve linking the abstract with the real-world phenomena and making decisions based on both. For others, mathematical literacy is that part of mathematics where the goal of mathematical activity is functional and alongside this, there is a part of mathematical activity where the goal is to explore and understand abstract structures and patterns for their own sake. Turner (2012) also discusses this question, as does Niss in Chap. 2 of this volume.

The PISA definition of mathematical literacy does not directly address this debate, and indeed the wording in the various definitions carefully steps around it. However, the definitions make it clear that mathematical literacy is the ability to use mathematical content (concepts, facts, procedures and tools) in real situations. It is also clear that teaching the school subject 'Mathematics' must address more than an 'abstract structures and skills' curriculum to develop students' mathematical literacy. Writing the preface to "Mathematics and Democracy" (Steen 2001), Orrill observes:

> An important theme of this volume, then, is that efforts to intensify attention to the traditional mathematics curriculum do not necessarily lead to increased competency with quantitative data and numbers. While perhaps surprising to many in the public, this conclusion follows from a simple recognition—that is, unlike mathematics, numeracy does not so much lead upward in an ascending pursuit of abstraction as it moves outward toward an ever richer engagement with life's diverse contexts and situations. When a professional mathematician is most fully at work, [the process becomes abstract]. The numerate individual, by contrast, seeks out the world and uses quantitative skills to come to grips with its varied settings and concrete particularity. (Orrill 2001, p. xviii)

Analysing Mathematical Literacy

The PISA Mathematics Framework defines mathematical literacy and the domain of mathematics for the PISA survey and describes the approach of the assessment. Figure 1.3 shows an overview of the main constructs of the 2012 Framework (OECD 2013a) and how they relate to each other.

The outer-most box in Fig. 1.3 shows that mathematical literacy is required to meet a challenge that arises in the real world. These challenges are categorised in two ways: by the nature of the situation (the context category) and the major domain of mathematics involved (the content category). The middle box highlights the nature of mathematical thought and action that needs to be used in solving this challenge. This is described in three ways: by mathematical content, by the fundamental mathematical capabilities that constitute mathematical activity and which are described in detail in Chaps. 2 and 4 of this volume (by Niss and by Turner, Blum and Niss respectively), and by the processes of mathematical modelling (discussed in detail in Chap. 3 of this volume by Stacey). The innermost box illustrates how the problem solver goes through these mathematical modelling processes in solving a problem.

A major purpose of the Framework is to specify the breadth of contexts, of mathematical thought and action and of solution processes that are included in the survey and the balance between them in the items. Figure 1.4 shows that there are six factors for which the Framework specifies the proportion of the items in the survey, relating to mode of assessment, content, context, process, response type, and difficulty (which is measured on a continuum rather than discretely). As well as being combined to make the overall score and ranking, three of these factors were separately reported for the 2012 survey: the continuing paper-based assessment and the new optional computer-based assessment (see below and also in Chap. 8), the content categories (four) and the processes (three). Reporting by process is a new feature of PISA 2012 that is discussed below. It has been introduced in order to give

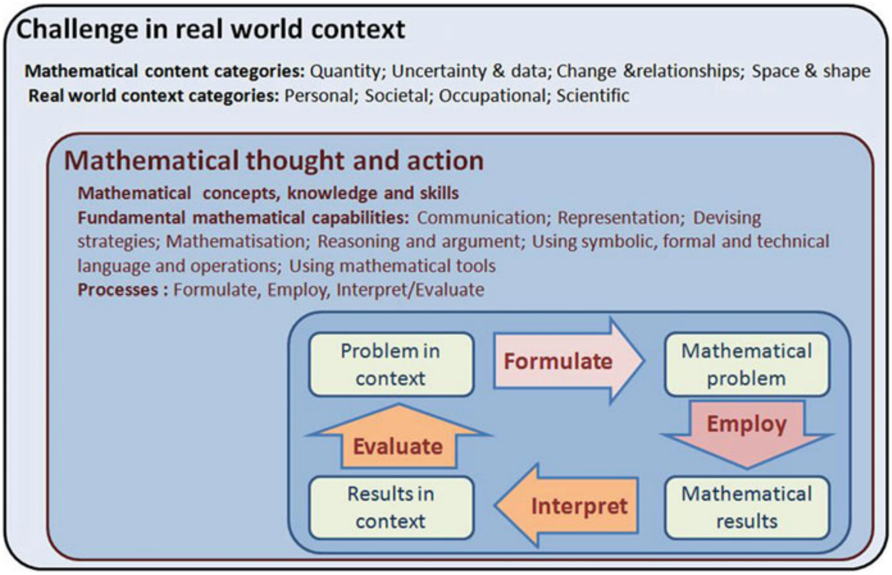


Fig. 1.3 A model of mathematical literacy in practice (OECD 2013a)

← Reporting categories →			← Further categories for balance →		
Assessment mode	**Process categories**	**Content categories**	**Context categories**	**Response Type**	**Cognitive Demand**
Paper-based Computer-based	Formulating situations mathematically Employing mathematical concepts, facts, procedures, and reasoning Interpreting, applying and evaluating mathematical outcomes	Quantity Uncertainty and data Change and Relationships Space and Shape	Personal Societal Occupational Scientific	Selected Response (multiple choice, complex multiple choice, variations) Constructed Response (expert, manual or auto-coded)	↑ ↑ Continuum of empirical difficulty ↓ ↓

Fig. 1.4 Categories over which the 2012 PISA Mathematics is reported and balanced

Table 1.1 Metadata for PM923 Sailing ships (PISA 2012 main survey)

	PM923 Sailing ships		
	Question 1	Question 3	Question 4
Assessment mode	Paper-based	Paper-based	Paper-based
Process category	Employ	Employ	Formulate
Content category	Quantity	Space and shape	Change and relationships
Context category	Scientific	Scientific	Scientific
Response format	Multiple choice	Multiple choice	Constructed response
Cognitive demand (item difficulty)	−0.9	−0.3	1.8

a better description of the abilities that underlie mathematical literacy. Previously PISA frameworks discussed processes by grouping items into 'competency classes/clusters' according to whether they required reproduction, connections or reflection, but outcomes were not reported using that classification. In Chap. 2 of this volume, Niss describes how the competency classes were linked to the other Framework elements. An early version of the competency classes as lower, middle and higher levels of assessment is found in de Lange (1992).

Each PISA item is classified according to these six factors, to ensure the balance of the assessment, and for aggregation of scores from designated items for reporting. Table 1.1 shows the relevant metadata for the three released items of PM923 Sailing ships. The assessment mode and response format are easily decided. The categorisations for process, content and context are determined by the Mathematics Expert Group and sometimes involve 'on balance' decisions. The cognitive demand is the item difficulty derived in advance of the main survey item selection from the field trial using Rasch-based item response theory (see, for example, Adams and Wu 2002). An average item has difficulty 0, items more difficult than average have positive scores and very difficult items have a score over 3.

Real-World Context Categories

Four context categories identify the broad areas of life from which the problem situations in the items may arise. For PISA 2012 these are labelled *Personal*, *Societal*, *Occupational* and *Scientific*. This is a simplification of the names for categories used in earlier PISA surveys, with minor adjustments of the scope of each. Formal definitions are given in the PISA 2012 Framework (OECD 2013a). Briefly, problems in a personal context arise from daily life with the perspective of the individual being central. Problems in a societal context arise from being a citizen, local, national, or global. Problems in an occupational context are from the world of work and problems in a scientific context (such as PM923 Sailing ships in Fig. 1.2) apply mathematical analysis to science and technology. From 2012, the *Scientific* category also includes problems entirely about mathematical constructs such as prime numbers (previously in the educational/occupational category), but because mathematical literacy is for functional use, extremely few PISA items are entirely intra-mathematical.

Earlier versions of the Framework described the different context categories as being of varying 'distance from the student' (with personal the closest, and scientific the furthest), which some observers criticised because of the great individual variation in students' experiences. This description was not used in 2012: instead the categories were effectively defined through multiple exemplifications. The four-way context categorisation is not rigorously defined, and can often be debated. Its only purpose is to ensure balance in the items of the PISA survey—they should arise from all the areas where mathematical literacy is important in order to fully represent the construct while engaging the interest of many types of students. The Framework specifies that about 25 % of the items should belong to each category.

Stacey's Chap. 3 of this volume addresses the contentious issue of selecting contexts for items that are authentic and relevant to students around the world, and Chaps. 6 (Turner) and 7 (Tout and Spithill) explain how this relevance is monitored by ratings from every participating country.

Content Categories

The outermost box of Fig. 1.3 shows that PISA problems are also categorised according to the nature of the mathematical phenomena that underlie the challenges and consequently the domains of mathematics that their solutions are likely to call upon. Starting from the 2003 survey, there have been four categories and approximately 25 % of items in the survey belong to each. The content categories of the PISA 2012 Framework (OECD 2013a) have previously been labelled 'big ideas' for PISA 2000 (OECD 1999) and 'overarching ideas' for the 2003, 2006 and 2009 surveys (OECD 2004, 2006, 2009c).

These content categories have a reasonable correspondence with divisions of the traditional school curriculum. So the items allocated to the content category *Quantity* tend to draw heavily on topics encountered under the headings of Number and Measurement, *Space and shape* items on Geometry, *Uncertainty and data* items on Probability and Statistics and *Change and relationships* on Algebra and Functions. However, the origin of the content categories is not from the school curriculum or from inside the discipline of mathematics. Instead, it reflects a movement towards phenomenological organisation that is intended to stress the underlying phenomena with which mathematics is concerned and to emphasise the unity of mathematics where ideas from different branches often work together to illuminate phenomena. Mathematical literacy tasks arising in real life often require mathematical concepts and procedures from various school or university topics to be used together. PISA items sometimes do.

It is also often the case that different good solutions can draw on different topics. For example PM923Q03 Sailing ships Question 3 could be solved by geometry and Pythagoras's theorem, but it might also be solved by making a scale drawing. PM977Q02 DVD Rental Question 2 (see Chap. 9 this volume or OECD 2013b) can be solved for full credit using either algebra or arithmetic reasoning. These difficulties are reduced by classifying PISA items on the underlying phenomenon that lies at the heart of the problem, rather than by the topic deemed by some expert to be appropriate.

As shown in Table 1.1, PM923Q01 Sailing ships Question 1 is categorised as *Quantity* because the essence is in the relative magnitude of the two wind speeds and the resultant percentage calculation. PM923Q03 Sailing ships Question 3 is categorised as *Space and shape* because of the geometric reasoning involved. PM923Q04 Sailing ships Question 4 is categorised as *Change and relationships* because the underlying challenge is to work with the savings as they increase over time. Because real-world challenges can involve many different thinking skills, on-balance decisions about where the main cognitive load arises often need to be made in this and other categorisations. As the PISA 2009 Framework explains:

> Each overarching idea represents a certain perspective or point of view and can be thought of as possessing a core, a centre of gravity, and somewhat blurred outskirts that allow for intersection with other overarching ideas. (OECD 2009a, p. 94)

Experience has shown that the four content categories, broadly interpreted, work well for an assessment of 15-year-olds. They provide sufficient variety and depth to reveal the essentials of mathematics and to stimulate the breadth that a good measure of mathematical literacy requires. They readily encompass the major problem types addressed within the compulsory years of school. It is frequently the case that more than one of the content categories is relevant to a proposed item, but it has never been the case that a potential item has been rejected because it cannot be placed within a content category. Theoretically, however, there is no claim that the four PISA content categories capture all of the phenomena that inspire mathematics. An exhaustive list would not be possible because of the breadth and variety of mathematics (OECD 2009b). As an example, the new phenomenon of 'information' as it applies to computer science and digital technology and modern biology (coding, security, transmission etc.) is now inspiring a great deal of mathematics but it is not clearly within any of the PISA content categories. However, experience has shown that the potential items (hence approachable by 15-year-olds) that have involved this phenomenon have had other characteristics that enable them to be placed within the current four categories. It is more usual for more than one of the content categories to be relevant to a proposed item than for none of them to be obviously relevant.

Behind Phenomenological Categorisation

The phenomenological organisation of mathematics has arisen in trying to identify unifying themes in the ever expanding and increasingly diversified discipline of mathematics. Steen (1990) edited a book that explored the 'developmental power' of five deep mathematical ideas (dimension, quantity, uncertainty, shape and change) relating to different types of pattern and which "nourish the growing branches of mathematics" (p. 3). He also identified other 'deep ideas' such as symmetry and visualisation, which recur in all parts of mathematics. Steen's five selected deep ideas have some commonality with the PISA content categories and indeed they are acknowledged as a source in the 2000, 2003, 2006 and 2009 frameworks.

PISA mathematics has also drawn inspiration from the Realistic Mathematics Education approach work of the famous Freudenthal Institute in the Netherlands, of which Jan de Lange, the Chair of the Mathematics Expert Group for the PISA 2000–2009, was a member (see, for example, de Lange 1987). Freudenthal (1991) saw mathematical concepts, structures, ideas and methods as serving to organise phenomena from the real world and from mathematics itself. For teaching, he valued problem situations that could be easily used by teachers to create in students the need to organise phenomena mathematically. Oldham (2006) explores these links.

There have been different approaches to describing mathematics from the problems that inspire it. For example, Bishop (1991) studied the mathematics of many different human cultures, aiming to identify universal characteristics. Because many cultures do not have a readily identifiable symbolic aspect to their mathematics, even the definition of mathematics is unclear, so deconstructing this is one of Bishop's aims. Bishop identified six 'environmental activities' (counting, measuring, locating, designing, playing and explaining) and claims that these are probably universal. Through many examples, he describes how these activities lead to the development of mathematics. In different cultures the end product mathematics may be different, but Bishop sees the commonality in the activities and the environmental needs that motivate them. Counting and measuring correspond broadly to PISA's *Quantity* and locating and designing correspond broadly to PISA's *Space and shape*. However, Bishop's description of playing links it to many underlying phenomena and he especially links explaining to classification and logic. Explaining in PISA fits better into the fundamental mathematical capabilities (see below). Bishop's cultural activity approach shares with PISA's phenomenological approach the intention to identify the human activities and concerns behind mathematics.

There are several consequences of PISA's decision to organise not around traditional curriculum topics but around the phenomena that inspire mathematics. One consequence, consistent with PISA's remit from the OECD to assess capacity to meet future challenges, is that there is no intention to systematically test a common core curriculum of participating countries as is done in TIMSS. Instead

PISA item writers begin by identifying problem situations that involve mathematical thinking. They aim for authentic situations, with obvious face validity, even if practical aspects of item presentation mean that considerable modification is needed. Tout and Spithill describe these processes in Chap. 7 of this volume.

Although PISA does not set out to test curriculum knowledge systematically, school curricula impinge strongly on the item writing and item selection process. An assessment of 15-year-olds must take into account the mathematics that they are likely to have learned, even though problems can often be solved without what teachers might think is the targeted knowledge. From a measurement perspective, it is useless to have items with only a tiny success rate. To this end, the PISA 2012 Framework, more than any of the earlier versions, includes a list of broadly described topics that might be required (e.g. 'linear and related equations and inequalities', 'basic aspects of the concept of probability'), supported by a survey of the mathematics standards for 11 high performing educational jurisdictions. This does not constitute a 'PISA curriculum' that is systematically tested, but it does guide item writers and gives participating countries better information about expected content. Topics do not belong to only one content category. Percentage calculations for example are likely to be common in problems inherently about quantity and also in problems about change. PM923 Sailing ships Questions 1 and 3 illustrate this (see Fig. 1.2 and Table 1.1). When the final item selection is being made, there are also checks to ensure that a good range of mathematical topics are involved, and that no particular mathematical skills are over-represented in the items. Turner in Chap. 6 of this volume describes such measures.

The Processes of Doing Mathematics

The mathematics frameworks for all PISA surveys have identified three key aspects of mathematical literacy items: the context and the content (as discussed above) and what is frequently called a 'process dimension' of mathematics—the activities that constitute doing and applying mathematics beyond Gee's (1998) 'design features' of mathematics. This dual nature of mathematics as content and process has long been widely recognised. For example, Georg Pólya (1962) who inspired much of the problem solving movement in mathematics education wrote:

> Our knowledge about any subject consists of *information* and of *know how*. If you have genuine *bona fide* experience of mathematical work on any level, elementary or advanced, there will be no doubt in your mind that, in mathematics, know-how is more important than mere possession of information. . . .What is know-how in mathematics? The ability to solve problems—not merely routine problems but problems requiring some degree of independence, judgment, originality, creativity. (p. vii)

The influential US report *Adding It Up* (Kilpatrick et al. 2001) lists five strands of mathematical proficiency: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning and productive disposition. The first two strands relate to mathematical content (Pólya's 'information') and the third and fourth describe mathematical process (Pólya's 'know-how'), and the fifth describes the intention to use these effectively (which is measured in PISA through the questionnaires). Around the world, there are many ways of describing this content-process distinction, and as is evident from Part III of this volume, PISA has been another prompt to highlight this.

Figure 1.3 depicts mathematical thought and action in three components. The first corresponds to the content aspect of mathematics (concepts, knowledge and skills) and the other two correspond to the process of solving problems with mathematics: the *fundamental mathematical capabilities* and the three '*processes*' of solving real problems. The fundamental mathematical capabilities describe mathematical actions that are involved in any mathematical activity, whilst the three processes refer to stages of action in solving real problems. Because of their centrality to the theorisation of PISA mathematics, they are each discussed below, and Chaps. 2 and 3 explore them in greater depth.

Fundamental Mathematical Capabilities

The PISA Frameworks described the *fundamental mathematical capabilities* differently over the years. This old idea is newly named for PISA 2012 to avoid conflicts within OECD material over the meaning of the previously used term ('competency'). They describe the type of activities that underlie any type of mathematical thought and problem solving. Abstract ideas have to be represented concretely (e.g. by a graphs or symbols), arguments have to be constructed, strategies for solving problems have to be described, calculations have to be carried out etc. The description of these mathematical thoughts and actions in PISA had its immediate roots in the work of Niss and colleagues in Denmark (Niss 1999; Niss and Højgaard 2011), who devised a set of eight competencies that together constitute mathematical competence. In Chap. 2, Niss gives a history of this development in Denmark and analyses how the Danish scheme was adopted and adapted in PISA. It also provides a useful guide to the confusing terminology changes that have beset this work. In Chap. 4 Turner, Blum and Niss describe how the fundamental mathematical capabilities can be used to describe the cognitive demand of mathematical tasks, in particular PISA items. They present empirical evidence that the difficulty of PISA items can in large part be predicted by analysing the items to see how deeply they call on each of the fundamental mathematical capabilities. As well as providing a useful tool for item and survey construction, understanding what contributes to increased demand for a capability can guide teachers towards what needs to be taught in mathematics, beyond just more content.

The fundamental mathematical capabilities cannot be individually assessed and reported by PISA, because from a psychometric point of view there are too many of them, and because they are rarely activated in isolation. Hence the 'process' aspect of mathematics is being reported for PISA 2012 through the more global *Formulate—Employ—Interpret* scheme that is described below. However, the work that

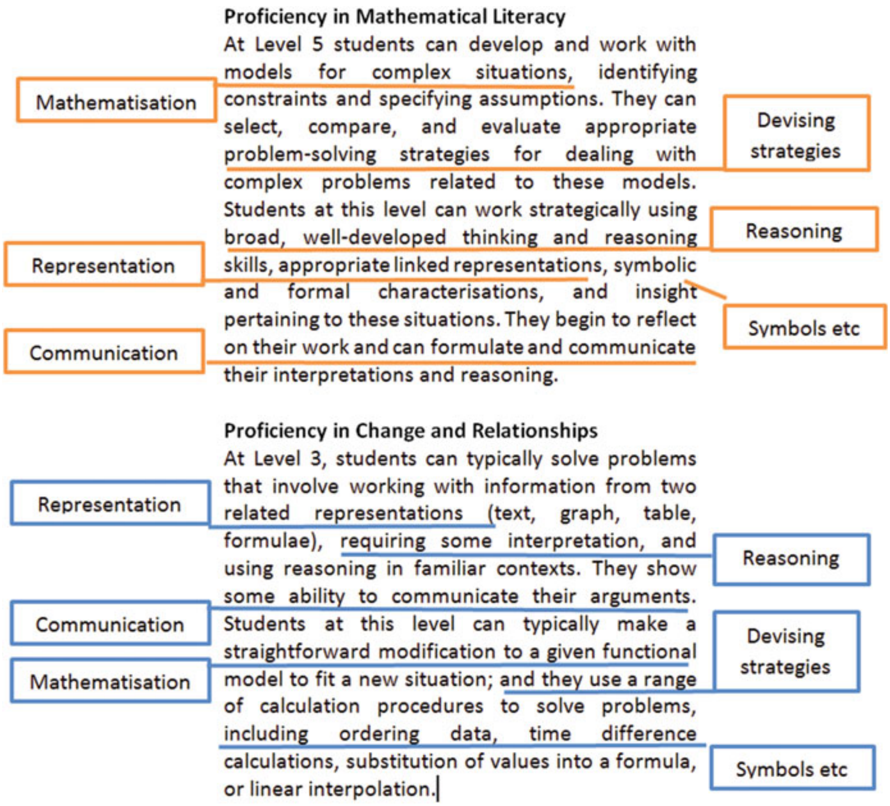


Fig. 1.5 Sample proficiency level descriptions showing references to fundamental mathematical capabilities

has been done in describing low and high level activation of the fundamental mathematical capabilities is the key to creating informative descriptions of the proficiency levels of students. Figure 1.5 shows two examples of how the fundamental mathematical capabilities appear in the description of proficiency—for overall proficiency for Level 5 and for the *Change and relationships* content category at Level 3 (OECD 2013d). The formal proficiency descriptions are given in the centre of the figure and the underlined sections point out the links to the capabilities. The different levels of activation of the capabilities become evident by comparing the descriptions across levels (see, for example OECD 2013d). In Chap. 4, Turner, Blum and Niss describe the increasing levels of activation of the capabilities from both theory driven and data driven approaches. Although they are not used formally for reporting or for balancing the item pool (as shown in Fig. 1.4), the fundamental mathematical capabilities are an essential feature of the Mathematics Framework and central to mathematical literacy.

Three Processes Linked to the Mathematical Modelling Cycle

Since 2000, PISA has reported by mathematics content categories, but reporting by the processes of mathematics is needed to provide countries with a full picture of the mathematical literacy of their students. This was an important innovation for PISA 2012 (OECD 2013a). The inner-most box of Fig. 1.3 portrays a model, idealised and simplified, of the stages through which a problem solver moves when exhibiting mathematical literacy. Mathematical literacy often begins with the "problem in context." The problem solver identifies the relevant mathematics in the problem situation, formulating the situation mathematically by imposing mathematical concepts, identifying relationships and making simplifying assumptions. This is the process of *Formulating situations mathematically*, abbreviated to '*Formulate*'. The problem solver has thus transformed the 'problem in context' into a mathematical problem, which is hopefully amenable to mathematical treatment. This is the process that both Sample Problem 4 (discussed earlier in this chapter) and PM923Q03 Sailing ships Question 3 (Fig. 1.2) involve and Sample Problem 3 (discussed earlier) does not.

The downward-pointing arrow in the inner-most box of Fig. 1.3 depicts the next process of *Employing mathematical concepts, facts, procedures, and reasoning* (abbreviated to *Employ*) to obtain mathematical results within the mathematical world of abstract objects. For Sailing ships Question 3, this is equivalent to solving Sample Problem 3. Next, the mathematical results are interpreted in terms of the original situation to obtain the 'results in context'. In the Sailing ships question, the numerical answer is interpreted as the length of the rope in metres. Furthermore, the adequacy of these results (and hence of the model) should be evaluated against the original problem. A serious solution of the Sailing ships question would need to take into account the precise purpose of solving the problem and consequently the required accuracy of the result. Is the amount of rope for tethering the kite at either end significant? Is it reasonable to assume the tethering rope lies in a plane? Does the deviation of the straight line model from a more accurate catenary matter? If it does matter, a new cycle of mathematical modelling may begin. In the context of a PISA assessment, these two stages have been combined to make one process *Interpreting, applying, and evaluating mathematical outcomes*, abbreviated to *Interpret*. This is because there are limited opportunities for any serious evaluation under the conditions of a PISA survey, in a short time by students sitting at a desk without additional resources. The key idea for PISA is to report separately on the two processes of moving between the real world and the mathematical world (*Formulate*, *Interpret*) and the process of working within the mathematical world (*Employ*).

In PM923 Sailing ships (Fig. 1.2), the judgement was made that for Question 1 and Question 3, the main demand was in carrying out the intra-mathematical work (see Table 1.1). PM923Q01 Question 1 requires a very small amount of formulation, discarding the extraneous information about 150 m height and seeing that the required quantity is 25 % more than the deck wind speed. Calculating this accurately is likely to be the major demand in this easy item. Similarly the intra-mathematical work is likely to be the most demanding aspect of PM923Q03 Question 3 (as discussed for Sample Problem 4 above). However for PM923Q04 Question 4, it is not the calculations, but identifying the relationships involved and how to put them together to build a solution that has been judged to be the most demanding aspect, and so Question 4 has been classified as *Formulate* in Table 1.1. Student performance on Question 4 is then pooled with performance on other items classified as *Formulate* to give a measure of proficiency on this process. Countries can use this measure to understand how well their students are learning to transform real problems into a form where mathematical analysis can be applied.

Just as the *fundamental mathematical capabilities* are used to describe overall proficiency, the degree of activation of them can be used to describe the levels of proficiency of students in the three processes. For example, among other capabilities, students who are at Level 4 of *Formulate* are described in the PISA 2012 report (OECD 2013d) as being able to link information and data from related representations (*representation* fundamental mathematical capability). This is higher activation than using only one representation.

The modelling cycle is a central aspect of the PISA conception of students as active problem solvers, and tasks that fully assess mathematical literacy will most probably involve all of these processes in the full modelling cycle. These are generally the favourite items of members of the Mathematics Expert Group. However, in the PISA survey it is important for the underlying psychometrics that students complete a large number of independent items in a short time. (Students in 2012 were presented with from 12 to 37 mathematics items, according to which particular booklets they were randomly assigned from the booklet rotation design.) Consequently, in most PISA items, the student is involved in only part of the modelling cycle. Items are classified according to the process that presents the highest demand for mathematical literacy within the item. This issue is explored in Chap. 3. In Chap. 11 in this volume, Ikeda discusses how tasks that focus on part of the modelling cycle can be an important part of teaching mathematical modelling. Of course over time, teaching must also give students extensive experience of tasks involving the whole modelling cycle.

The Framework specifies that about half of the mathematics items used in the PISA survey are classified as *Employ* and about one quarter are in each of the *Formulate* and *Interpret* categories. Mathematisation and the mathematical modelling cycle have always had a substantial role in the PISA frameworks but 2012 was the first survey to report results according to the modelling cycle processes. Because of its centrality to PISA, Chap. 3 of this volume by Stacey discusses the theoretical background and practical considerations of the assessment of mathematics as applied in the real world.

Computer-Based Assessment of Mathematics

For the first time, PISA 2012 supplemented the paper-based assessment with an optional computer-based assessment of mathematical literacy, abbreviated to CBAM. In 2012, 32 countries took up this option. This follows two earlier PISA initiatives: the computer-based assessment of science beginning in 2006 and a digital reading assessment beginning in 2009. CBAM items are presented on a computer, students respond on the computer, and they can also use pencil and paper to assist their thinking. In Chap. 7 of this volume Tout and Spithill describe the development of CBAM items, and in Chap. 8 Bardini analyses their characteristics.

Computer technology can alter assessment from the points of view of the student and the assessor. It can alter all phases of assessment: how tasks are selected (e.g. they might be automatically generated from an item pool), how they are presented, how students should operate while responding and with what tools, how the evidence provided by students is identified, and how this evidence is accumulated across tasks (Almond et al. 2003). The review by Stacey and Wiliam (2013) provides a wide range of examples of fruitful directions for these potential improvements, which range from simple changes in items to assessment of authentic tasks by collaborative groups in virtual environments. For PISA, these changes are just beginning.

In 2015, students in most countries will take the PISA mathematical literacy assessment at a computer. Items previously used in the paper-based survey will be presented on computer (OECD 2013c). An equivalent paper-based assessment will be used in countries without adequate infrastructure in schools. The advantages anticipated from this approach stem from simplified survey administration and greatly simplified processing of survey responses. The intention is that as far as possible the measure of mathematical literacy remains comparable with that of previous paper-based surveys despite the change in delivery mode (and this will be monitored).

Importantly, CBAM in PISA 2012 had a different philosophy. Just as the PISA Digital Reading survey was a response to the observation that in all walks of life, citizens now use digital resources to obtain information and communicate with friends and businesses, CBAM was a response to the changing face of mathematical literacy in a technology-rich world, where computerisation is rapidly changing the face of occupational, social and personal life (Frey and Osborne 2013). Consequently, a main task of the 2012 Mathematics Framework development was to define the new proficiency to be assessed by CBAM. What should the items and the assessment process be like?

Computer technology provides a communications infrastructure as well as a substantial computational infrastructure. Technology can support remarkable changes in the presentation of items and in the way students operate on them. It can provide simple computational aids (such as the many online calculators that abound on commercial websites) or it can provide open computational tools of remarkable power, including spreadsheets, function graphing, statistical software and computer algebra systems. The 2012 Framework embraced all of these aspects as theoretically part of CBAM.

One function of CBAM was recognised as enhancing the assessment of 'traditional' mathematical literacy beyond what can be achieved with a paper-based assessment. In this function, computer-based assessment can extend the range of phenomena that inspire viable PISA items, for example by using a dynamic stimulus for an item involving movement or by providing a rotatable three dimensional image to mimic the way in which a real object can be handled, or by realistically including modern-day website interactions. By having enhanced visual presentation and action responses, computer-based assessment may incidentally reduce the influence of verbal ability on mathematics scores. Chapters 7 and 8 give many examples.

The second function of the Framework analysis for CBAM was to demonstrate how mathematical literacy may itself be changing in a computationally rich world. This required considerations of changes in the workplace as well as changes in mathematical practice. The impact of computer technology on the ways in which individuals use mathematics, and consequently should learn it, has long been discussed and continues to evolve. Over the previous 40 years, the practical importance of pen and paper arithmetic algorithms has withered to close to zero, being gradually replaced by mental computation and estimation when feasible, backed up by computer or calculator use (Cockcroft 1982). This trend is accelerating, and applying now to mathematical routines across all topics (e.g. algebra, statistics, data presentation, functions) not just basic arithmetic.

The explicit mathematics of computation is increasingly embedded in the tools we use and consequently is increasingly invisible in people's lives. Shopping provides a daily reminder. It is no longer the shop assistant but the computerised technology at the cash register that weighs vegetables, multiplies weights by unit costs to get the prices, adds them up to get the bill and subtracts to calculate the change. The computer takes over the computational load so that what many people regard as 'the mathematics' is no longer evident.

Changes in mathematics in the workplace go beyond this. It is not just that the shop assistant no longer works out the bill. Behind the scenes, the shop manager has access to a vast web of data on purchasing and products. This needs to be insightfully utilised to run a business effectively. We now live increasingly in a society "drenched in data" (Steen 1999, p. 9) where

> computers meticulously and relentlessly note details about the world around them and carefully record these details. As a result, they create data in increasing amounts every time a purchase is made, a poll is taken, a disease is diagnosed, or a satellite passes over a section of terrain. (Orrill 2001, p. xvi)

Handling these large data sets or their automatically generated summary data (e.g. in control systems) and interacting flexibly and intelligently with them will increasingly become a common stimulus for employing mathematical literacy. The National Research Council's study of massive data analysis (2013) points to the technical challenges but it also points to the centrality of inference, having people who can turn data into knowledge. Sound inference is an aspect of mathematical literacy, with or without computer-based assessment. In the introduction, the National Research Council (2013) observes that there were six fields strongly affected by massive data analysis at the turn of the century when the quotes above by Steen and Orrill were written, and thirteen strongly affected by 2012.

After analysing the mathematical literacy required in industry and business to respond to the new data-rich, visualisation-rich and computationally-rich environment, Hoyles et al. (2010) coined the term 'techno-mathematical literacies' to describe the inter-dependence of mathematical literacy and the use of information technology for employees at all levels in the workplace. In responding to computer-based items, students encounter cognitive demands from three sources:

- from using the technology itself (e.g. using a mouse, knowing computer conventions such as the back button for moving around websites)
- from mathematical literacy inherent in the problem independent of technology
- from the techno-mathematical literacies at the interface of mathematics and technology.

The intention is that the first of these should be minimised, the second is familiar and the third is rapidly becoming part of mathematical literacy. Using specialised workplace systems and also open mathematical tools, especially for statistics, graphing, data handling, three dimensional visualisation and algebra requires both the understanding of the underlying mathematics as well as being able to think in the ways that using the technology demands. Some of the challenges and opportunities in assessing mathematics supported by such tools are reviewed by Stacey and Wiliam (2013). CBAM is an expansion of existing policies of PISA mathematics has had for allowing calculator use: students should make sensible choices to use or not to use their tools as the problem requires; it is not calculator use itself that is tested. Despite all the changes and new opportunities afforded by the increasing use of computers including in the assessment context, it was judged that the major categorisations of the items, as shown in Fig. 1.4, could be taken across to apply to CBAM, without major change.

The optional CBAM of 2012 was a small first step towards developing the new assessment, constrained by both the complexity of delivering an untried component and the likely abilities of students around the world in this new area. However, it was an important step. Full participation in society and in the workplace in this information-rich world requires an expanded view of mathematical literacy.

Surveying Attitudes and School Context

In addition to measuring students' mathematical literacy, PISA uses questionnaires for students and schools to measure the attitudes towards learning that are likely to make them successful life-long learners and to gather information that can help explain what promotes good outcomes of schooling. Attitudes and emotions (e.g. self-confidence, curiosity) are not defined as components of mathematical literacy. This contrasts with some frameworks that are focused on teaching. For example, Kilpatrick et al. (2001) identify 'productive dispositions' as one of the strands on mathematical proficiency. PISA does not include these personal qualities as part of mathematical literacy, but recognises that it is unlikely that students who do not exhibit productive dispositions will develop their mathematical literacy to the full (OECD 2006).

For mathematics, the Context Questionnaire Framework for PISA 2012 specifies "information about students' experience with mathematics in and out of school [. . .], motivation, interest in mathematics and engagement with mathematics" as well as aspects of learning and instruction, learning and teaching strategies and links to school structures and organisation (OECD 2013a, p. 182). Questions about motivation and intentions to work hard and to continue with the study of mathematics at school are seen as especially important, not just because there is a positive correlation between attitudes and performance, but also because of the concern by governments around the world to boost the STEM workforce (science, technology, engineering and mathematics). The PISA 2012 Framework (OECD 2013a) provides the reasons behind the choices of questionnaire themes and items. In Chap. 10 of this volume, Cogan and Schmidt describe one of the most interesting aspects for PISA 2012, the innovative investigation of opportunity to learn with specific regard to items varying on dimensions relevant to mathematical literacy.

Conclusion

This chapter has offered an introduction to the assessment frameworks for the first several PISA surveys and their key concepts, and given insight into the underpinning ideas and some of the related scholarship that have influenced the Mathematics Expert Groups from 2000 to 2012 in framework development. Preparation of the framework involves two main tasks: to clearly define the domain that is to be assessed, and to analyse the domain so that the resulting item set provides comprehensive coverage of the domain from multiple points of view and so that descriptions of students' increasing proficiency reveal the fundamental capabilities that contribute to success.

It is perhaps worth explicitly noting that decisions made in an assessment framework really affect the results of that assessment. Making different choices of what to assess, or choosing a different balance of the items in various categories makes a difference in all outcomes, including international rankings. One illustration of this is that the two major international surveys of mathematics, PISA and Trends in Mathematics and Science Study (TIMSS) produce different international rankings. In contrast to PISA's focus on mathematical literacy, TIMSS begins with a thorough analysis of the intended school curricula of participating countries and designs items to test this (Mullis et al. 2009). The systematic differences in results have been analysed in many publications (e.g. Wu 2010). Within the PISA approach, changing the proportions of items in each Framework category would also change results, because countries vary in their performance across categories. For these theoretical and practical reasons, the choices made in devising the PISA Frameworks matter.

As outlined above, there have been many changes in the Mathematics Frameworks but this is best seen as a process of evolution in response to feedback from many sources, rather than revolution. The core idea of mathematical literacy has been strongly held through the 2000–2012 surveys, extended now to encompass the new directions that arise as the personal, societal, occupational and scientific environment is gradually transformed by technology.

References

Adams, R., & Wu, M. (Eds.). (2002). *The PISA 2000 technical report*. Paris: OECD Publications.

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2003). *A four-process architecture for assessment delivery, with connections to assessment design*. Los Angeles: University of California Los Angeles Center for Research on Evaluations, Standards and Student Testing (CRESST).

Autor, D., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics, 118*(4), 1279–1334.

Bishop, A. (1991). *Mathematical enculturation: A cultural perspective on mathematics education*. Dordrecht: Kluwer.

Boyer, C. (1968). *A history of mathematics*. New York: Wiley.

Bybee, R. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth: Heinemann.

Cockcroft, W. (1982). *Mathematics counts. Report of the committee of inquiry into the teaching of mathematics in schools under the chairmanship of Dr W. H. Cockcroft*. London: Her Majesty's Stationery Office. http://www.educationengland.org.uk/documents/cockcroft/cockcroft1982.html#03. Accessed 10 Nov 2013.

Comber, B. (2013). Critical theory and literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–10). Oxford: Blackwell Publishing Ltd. doi:10.1002/9781405198431.wbeal0287.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37* (6), 582–601.

de Lange, J. (1987). *Mathematics—Insight and meaning*. Utrecht: Rijksuniversiteit Utrecht.

de Lange, J. (1992). No change without problems. In M. Stephens & J. Izard (Eds.), *Reshaping assessment practices: Assessment in the mathematical sciences under challenge. Proceedings from the first national conference on assessment in the mathematical sciences* (pp. 46–76). Melbourne: ACER.

de Lange, J. (2006). Mathematical literacy for living from OECD-PISA perspective. *Tsukuba Journal of Educational Study in Mathematics, 25*, 13–35. http://www.criced.tsukuba.ac.jp/math/apec2006/Tsukuba_Journal_25.pdf

Freudenthal, H. (1991). *Revisiting mathematics education—China lectures*. Dordrecht: Kluwer.

Frey, C. B., & Osborne, M. A. (2013). The future of employment: How susceptible are jobs to computerisation? Oxford Martin School working paper. http://www.futuretech.ox.ac.uk/sites/futuretech.ox.ac.uk/files/The_Future_of_Employment_OMS_Working_Paper_0.pdf. Accessed 29 Sept 2013.

Gee, J. (1998). *Preamble to a literacy program*. Madison: Department of Curriculum and Instruction.

Hoyles, C., Noss, R., Kent, P., & Bakker, A. (2010). *Improving mathematics at work: The need for techno-mathematical literacies*. London: Routledge.

Hughes-Hallett, D. (2001). Achieving numeracy: The challenge of implementation. In L. A. Steen (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. 93–98). Princeton: National Council on Education and the Disciplines. http://www.maa.org/sites/default/files/pdf/QL/MathAndDemocracy.pdf. Accessed 28 Sept 2013.

Keskpaik, S., & Salles, F. (2013). *Note d'information. Les élèves de 15 ans en France selon PISA 2012 en culture mathématique : baisse des performances et augmentation des inégalités depuis 2003*. Paris: DEPP B2 Ministère éducation nationale.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Boston: TIMSS & PIRLS International Study Center.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston: Author.

National Research Council. (2013). *Frontiers in massive data analysis*. Washington, DC: National Academies Press.

Niss, M. (1999). *Kompetencer og uddannelsesbeskrivelse* [Competencies and subject description]. *Uddannelse, 9*, 21–29.

Niss, M., & Højgaard, T. (Eds.). (2011). *Competencies and mathematical learning: Ideas and inspiration for the development of teaching and learning in Denmark* (IMFUFA tekst). Roskilde: Roskilde University.

Oldham, E. (2006). The PISA mathematics results in context. *The Irish Journal of Education (Iris Eireannach an Oideachais), 37*, 27–52.

Organisation for Economic Co-operation and Development (OECD). (1995). *Literacy, economy, and society: Results of the first International Adult Literacy Survey*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris: PISA, OECD Publishing. doi:10.1787/9789264181564-en.

Organisation for Economic Co-operation and Development (OECD). (2004). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: PISA, OECD Publishing. doi:10.1787/9789264101739-en.

Organisation for Economic Co-operation and Development (OECD). (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: PISA, OECD Publishing. doi:10.1787/9789264026407-en.

Organisation for Economic Co-operation and Development (OECD). (2009a). *PISA: Take the test*. Paris: OECD Publications. http://www.oecd.org/pisa/pisaproducts/Take%20the%20test%20e%20book.pdf. Accessed 17 May 2014.

Organisation for Economic Co-operation and Development (OECD). (2009b). *Learning mathematics for life: A perspective from PISA*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2009c). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD Publications. doi:10.1787/9789264062658-en.

Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. doi:10.1787/9789264190511-en.

Organisation for Economic Co-operation and Development (OECD). (2013b). *PISA 2012 released mathematics items*. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf. Accessed 8 Oct 2013.

Organisation for Economic Co-operation and Development (OECD). (2013c). *PISA 2015 draft mathematics framework*. Paris: OECD. http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Mathematics%20Framework%20.pdf. Accessed 25 Oct 2013.

Organisation for Economic Co-operation and Development (OECD). (2013d). *PISA 2012 results: What students know and can do (Vol. I): Student performance in mathematics, reading and science*. Paris: PISA, OECD Publishing. doi:10.1787/9789264201118-en.

Orrill, R. (2001). Mathematics, numeracy and democracy. In L. A. Steen (Ed.), *Mathematics and democracy: The case for quantitative literacy*. New Jersey: NCED.

Pólya, G. (1962). *Mathematical discovery. On understanding, learning and teaching problem solving*. New York: Wiley.

Stacey, K. (in press). The international assessment of mathematical literacy: PISA 2012. In S. -J. Cho (Ed.), *Selected regular lectures from the 12th international congress of mathematical education.* Seoul: International Committee on Mathematical Instruction (ICMI).

Stacey, K., & Wiliam, D. (2013). Technology and assessment in mathematics. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 721–752). Springer: New York.

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

Steen, L. A. (1999). Numeracy: The new literacy for a data-drenched society. *Educational Leadership, 57*(2), 8–13.

Steen, L. A. (Ed.). (2001). *Mathematics and democracy: The case for quantitative literacy*. New Jersey: National Council on Education and the Disciplines. http://www.maa.org/sites/default/files/pdf/QL/MathAndDemocracy.pdf. Accessed 28 Sept 2013.

Turner, R. (2012). *Mathematical literacy: Are we there yet?* Paper presented at ICME-12, Topic Study Group 6: Mathematics literacy. Seoul, July 2012. http://works.bepress.com/ross_turner/22. Accessed 24 Nov 2013.

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.

Chapter 2
Mathematical Competencies and PISA

Mogens Niss

Abstract The focus of this chapter is on the notion of mathematical competence and its varying role in the PISA mathematics frameworks and reports of PISA results throughout the first five survey administrations, in which mathematical literacy is a key concept. The chapter presents the genesis and development of the competency notion in Denmark, especially in the so-called KOM project, with a view to similar or related notions developed in different environments and contexts, and provides a detailed description of the eight competencies identified in the KOM project. Also the relationship between the mathematical competencies and the fundamental mathematical capabilities of the PISA 2012 Framework is outlined and discussed.

Introduction

The notion of mathematical competence—which will be introduced and discussed in greater detail below—has been present in some way or another in all the PISA mathematics frameworks from the very beginning in the late 1990s. However, the actual role of mathematical competencies in the PISA frameworks and in the reporting of PISA outcomes has been subject to considerable evolution across the five PISA surveys completed so far; that is, until 2013.

These facts provide sufficient reason for including a chapter on the role of mathematical competencies within PISA in this book. The structure of the chapter is as follows. After this introduction comes a section in which the genesis of the notion of mathematical competence is presented and its history briefly outlined. It may be worth noticing that the inception of this notion—in the specific version presented in this chapter—took place more or less at the same time but completely independently of the launching of PISA in 1997. Subsequently, the trajectories of development of mathematical competencies and PISA, respectively, became intertwined in several interesting ways. The section to follow next considers further

M. Niss (✉)
IMFUFA/NSM, Roskilde University, Universitetsvej 1, Bldg. 27, 4000, Roskilde, Denmark
e-mail: mn@ruc.dk

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_2

aspects of the notion of mathematical competence in a general setting not specifically focused on PISA. Then comes the core of this chapter, namely an analysis and discussion of the changing role of mathematical competencies within PISA, both in relation to the mathematics frameworks of the different PISA survey administrations, and to the reporting of PISA outcomes. That section also includes a discussion of the transformation of the original competencies into a particular set of competencies that have proved significant in capturing and characterising the intrinsic demands of PISA items.

Brief History of the General Notion of Competencies and a Side View to Its Relatives

Traditionally, in most places mathematics teaching and learning have been defined in terms of a curriculum to be taught by the teacher and achieved by the student. Typically, a curriculum used to be a sequence—organised by conceptual and logical progression—of mathematical concepts, terms, topics, results and methods that people should know, supplemented with a list of procedural and technical skills they should possess. In curriculum documents, the generally formulated requirements are often accompanied by illustrative examples of tasks (including exercises and problems) that students are expected to be able to handle when/if they have achieved the curriculum.

However, there have always been mathematics educators (e.g. Hans Freudenthal (1973, 1991), who kept emphasising that mathematics should be perceived as an activity) who have insisted that coming to grips with what it means to be mathematically competent cannot be adequately captured by way of such lists. There is significantly more to be said, they believe, in the same way as no sensible person would reduce the definition of linguistic competence in a given language to lists of the words, orthography and grammatical rules that people have to know in that language. Already in the first IEA study (Husén 1967), the precursor to and generator of the later TIMSS studies, mathematics is defined by way of two dimensions, mathematical topics and five cognitive behaviour levels:

> (a) knowledge and information: recall of definitions, notations, concepts; (b) techniques and skills: solutions; (c) translation of data into symbols or schemas and vice versa; (d) comprehension: capacity to analyse problems, to follow reasoning; (e) inventiveness: reasoning creatively in mathematics. (Niss et al. 2013, p. 986)

Heinrich Winter (1995) spoke about three fundamental, general experiences that mathematics education should bring about: coming to grips with essential phenomena in nature, society and culture; understanding mathematical objects and relations as represented in languages, symbols, pictures and formulae; fostering the ability to engage in problem solving, including heuristics.

Also, the notions of numeracy, mathematical literacy, and quantitative literacy have been coined so as to point to essential features of mathematical mastery, geared towards the functional use of mathematics, that go beyond factual knowledge and procedural skills (see also Chap. 1 in this volume). Moreover, newer curriculum documents such as the NCTM Standards of 1989 (National Council of Teachers of Mathematics 1989) also involve components that are not defined in terms of factual knowledge and procedural skills. The *Standards* identify five ability-oriented goals for all K-12 students: (1) that they learn to value mathematics, (2) that they become confident in their ability to do mathematics, (3) that they become mathematical problem solvers, (4) that they learn to communicate mathematically, and (5) that they learn to reason mathematically (NCTM 1989, p. 5).

Let these few examples suffice to indicate that lines of thought do exist that point to (varying) aspects of mathematical mastery that go beyond content knowledge and procedural skills. The notion of mathematical competence and competencies was coined and developed in the same spirit, albeit not restricted to functional aspects as above.

From the very beginning, the graduate and undergraduate mathematics studies at Roskilde University, Denmark, designed and established in 1972–1974, and continuously developed since then, were described partly in terms of the kinds of overarching mathematical insights and competencies (although slightly different words were used at that time) that graduates were supposed to develop and possess upon graduation. Needless to say, the programme documents also included a list of traditional mathematical topics that students should become familiar with. For a brief introduction to the mathematics studies at Roskilde University, see Niss (2001). In the 1970s and 1980s aspects of this way of thinking provided inspiration for curriculum development in lower and upper secondary mathematics education in Denmark.

In the second half of the 1990s executives of the Danish Ministry of Education wanted the Ministry to chart, for each school subject, what was called ‘the added value’ generated within the subject by moving up through the education levels, from elementary and primary (Grades K-6), over lower secondary (Grades 7–9) through to the upper secondary levels (Grades 10–12 in different streams), with a special emphasis on the latter levels. It was immediately clear to the mathematics inspectors and other key officers in the Ministry that the added value could not be determined in a sensible manner by merely pointing to the new mathematical concepts, topics and results that are put on the agenda in the transition from one level or grade to the next. But what else could be done? The officers in the Ministry turned to me for assistance, and after a couple of meetings I devised a first draft of what eventually became a system of mathematical competencies. The underlying thinking was greatly influenced by the philosophy underpinning the mathematics studies at Roskilde University. The fundamental idea was to try to answer two questions.

The first question springs from noting that any observer of mathematics teaching and learning in the education system, at least in Denmark, will find that what happens in elementary and primary mathematics classrooms, in lower secondary classrooms, in upper secondary classrooms and, even more so, in tertiary classrooms, displays a dramatic variability, not only because the mathematical topics and concepts dealt with in these classrooms are different, but also, and more importantly, because topics, concepts, questions and claims are dealt with in very different ways at different levels—in particular when it comes to justification of statements—even to the point where mathematics at, say, the primary level and at the tertiary level appears to be completely different subjects. So, given this variability, what is it that makes it reasonable to use the same label, mathematics, for all the different versions of the subject called mathematics across education levels? Differently put, what do all these versions have in common, apart from the label itself? Next, if we can come up with a satisfactory answer to the question of what very different versions of mathematics have *in common*, the second question is then to look into how we can use this answer to account, in a unified and non-superficial manner, for the obvious *differences* encountered in mathematics education across levels.

As we have seen, the commonalities in the different versions of mathematics do not lie in any *specific* content, as this is clearly very different at different levels. Whilst it is true that content introduced at one level remains pertinent and relevant for all subsequent levels, new content is introduced at every level. The general rational numbers of the lower secondary level are not dealt with at the primary level. The trigonometry or the polynomials of the upper secondary level have no presence at the primary or lower secondary levels. The general vector spaces, analytic functions or commutative rings of the tertiary level have no place at the upper secondary level. In other words, in terms of specific content, the only content that is common to all levels are small natural numbers (with place value) and names of well-known geometrical figures. Well, but instead of specific content we might focus on more abstract *generic* content such as numbers and the rules that govern them, geometric figures and their properties, measure and mensuration, all of which are present at any level, albeit in different manifestations. Yes, but the intersection would still be very small, as a lot of post-elementary mathematics cannot be subsumed under those content categories. Of course, we might go further and adopt a *meta-perspective* on content, as is done in PISA, and consider phenomenological content categories such as *Space and shape*, *Change and relationships*, *Quantity*, and *Uncertainty and data*, all of which are present at any level of mathematics education. However, this does not in any way imply that these categories cover all mathematical content beyond the lower secondary level. For example, an unreasonable amount of abstraction and flexibility of interpretation would be required to fit topics such as integration, topological groups or functional analysis into these categories. Finally, one might consider taking several steps up the abstraction ladder and speak, for example, of mathematics as a whole as the science of patterns (Devlin 1994, 2000), a view that does provide food for thought but is also so abstract and general that one may be in doubt of what is actually being said and covered by that statement. If, for instance, people in chemistry, in botany, or in art and design wished to claim—which wouldn't seem unreasonable—that they certainly profess sciences of patterns, would we then consider these sciences part of mathematics? Probably not.

Instead of focusing on content, I chose to focus on mathematical activity by asking what it means to be mathematically competent. What are the characteristics of a person who, on the basis of knowledge and insight, is able to successfully deal with a wide variety of situations that contain explicit or implicit mathematical challenges? *Mathematical competence* is the term chosen to denote this aggregate and complex entity. I wanted the answers to these questions to be specific to mathematics, even if cast in a terminology that may seem generalisable to other subjects, to cover all age groups and education levels, and to make sense across all mathematical topics, without being so general that the substance evaporates. The analogy with linguistic competence touched upon above was carried further as an inspiration to answering these questions. If linguistic competence in a language amounts to being able to understand and interpret others' written texts and oral statements and narratives in that language, as well as to being able to express oneself in writing and in speech, all of this in a variety of contexts, genres and registers, what would be the counterparts with regard to mathematics? Clearly, people listen, read, speak and write about very different things and in very different ways when going to kindergarten and when teaching, say, English history to PhD students. However, the same four components—which we might agree to call linguistic competencies—play key parts at all levels.

Inspired by these considerations, the task was to identify the key components, the *mathematical competencies* analogous to linguistic competencies, in *mathematical competence*. The approach taken was to reflect on and theoretically analyse the mathematical activities involved in dealing with mathematics-laden, challenging situations, taking introspection and observation of students at work as my point of departure.

It is a characteristic of mathematics-laden situations that they contain or can give rise to actual and potential *questions*—which may not yet have been articulated—to which we seek valid answers. So, it seems natural to focus on the competencies involved in posing and answering different sorts of questions pertinent to mathematics in different settings, contexts and situations. The first competency then is to do with key aspects of mathematical thinking, namely the nature and kinds of questions that are typical of mathematics, and the nature and kinds of answers that may typically be obtained. This is closely related to the types, scopes and ranges of the statements found in mathematics, and to the extension of the concepts involved in these statements, e.g. when the term 'number' sometimes refers to natural numbers, sometimes to rational numbers or complex numbers. The ability to relate to and deal with such issues was called the *mathematical thinking competency*. The second competency is to do with identifying, posing and solving mathematical problems. Not surprisingly, this was called the *mathematical problem handling competency*. It is part of the view of mathematics education nurtured in most places in Denmark, and especially at Roskilde University, that the place and role of mathematics in other academic or practical domains are crucial to mathematics education. As the involvement of mathematics in extra-mathematical domains takes place by way of explicit or implicit mathematical models and modelling, individuals' ability to deal with existing models and to engage in model construction (active modelling) is identified as a third independent competency, the *mathematical modelling competency*. The fourth and last of this group of competencies focuses on the ways in which mathematical claims, answers and solutions are validated and justified by mathematical reasoning. The ability to follow such reasoning as well as to construct chains of arguments so as to justify claims, answers and solutions was called the *mathematical reasoning competency*.

The activation of each of these four competencies requires the ability to deal with and utilise mathematical language and tools. Amongst these, various representations of mathematical entities (i.e. objects, phenomena, relations, processes, and situations) are of key significance. Typical examples of mathematical representations take the form of symbols, graphs, diagrams, charts, tables, and verbal descriptions of entities. The ability to interpret and employ as well as to translate between such representations, whilst being aware of the sort and amount of information contained in each representation, was called the *mathematical representation competency*. One of the most important categories of mathematical representations consists of mathematical symbols, and expressions composed of symbols. The ability to deal with mathematical symbolism—i.e. symbols, symbolic expressions, and the rules that govern the manipulation of them—and related formalisms, i.e. specific rule-based mathematical systems making extensive use of symbolic expressions, e.g. matrix algebra, was called the *mathematical symbols and formalism competency*. Considering the fact that anyone who is learning or practising mathematics has to be engaged, in some way or another, in receptive or constructive communication about matters mathematical, either by attempting to grasp others' written, oral, figurative or gestural mathematical communication or by actively expressing oneself to others through various means, a *mathematical communication competency* is important to include. Finally, mathematics has always, today as in the past, made use of a variety of physical objects, instruments or machinery, to represent mathematical entities or to assist in carrying out mathematical processes. Counting stones (calculi), abaci, rulers, compasses, slide rulers, protractors, drawing instruments, tables, calculators and computers, are just a few examples. The ability to handle such physical aids and tools (mathematical technology in a broad sense) with insight into their properties and limitations is an essential competency of contemporary relevance, which was called the *mathematical aids and tools competency*. In the next section, a figure depicting the competencies as the petals of a flower is presented (Fig. 2.1).

We now have identified eight mathematical competencies, which are claimed to form an exhaustive set of constituents of what has been termed mathematical competence. The first published version of these competencies (in Danish) can be found in Niss (1999) in a journal published then by the Danish Ministry of Education. Each of the competencies can be perceived as the ability to successfully deal with a wide variety of situations in which explicit or implicit mathematical challenges of a certain type manifest themselves. By addressing and playing out in mathematics-laden situations, the competencies do not deal with mathematics as a whole. Therefore, the set of competencies was complemented with three kinds of *overview and judgement concerning mathematics as a discipline*: the actual use of

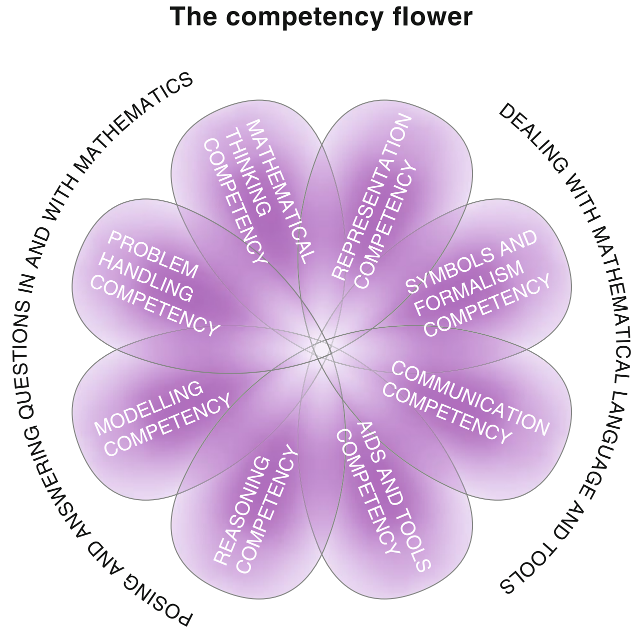


Fig. 2.1 The 'competency flower' from the KOM project

mathematics in society and in extra-mathematical domains, the specific nature and characteristics of mathematics as a discipline compared and contrasted with other scientific and scholarly disciplines, and the historical development of mathematics in society and culture.

Soon after, in 2000, a Danish government agency and the Danish Ministry of Education jointly established a task force to undertake a project to analyse the state of affairs concerning the teaching and learning of mathematics at all levels of the Danish education system, to identify major problems and challenges within this system, especially regarding progression of teaching and learning and the transition between the main sections of the system, and to propose ways to counteract, and possibly solve, the problems thus identified. I was appointed director of the project with Tomas Højgaard Jensen serving as its academic secretary. The project became known as the *KOM project* (KOM = Kompetencer og matematiklæring, in Danish, which means "Competencies and the learning of mathematics"), because the main theoretical tool adopted by the task force to analyse mathematics education in Denmark was the set of eight mathematical competencies, and the three kinds of overview and judgement, introduced above. More specifically, the actual presence and role of the various competencies in mathematics teaching and learning at different levels were analysed. This allowed for the detection of significant differences in the emphases placed on the individual competencies in different sections of the education system. This in turn helped explain some of the observed problems of transition between the sections as well as insufficient progression of teaching and learning within the entire system. The competencies were also used in a normative manner to propose curriculum designs, modes and instruments of assessment, and competency-oriented teaching and learning activities from school to university, teacher education included. In the next section we shall provide a more detailed account of further aspects of the competencies and their relationship with mathematical content. The formal outcome of the KOM project was the publication, in Danish, of the official report on the project (Niss and Jensen 2002). However, during and after the completion of the project a huge number of meetings, seminars and in-service courses were held throughout Denmark and in other countries to disseminate and discuss the ideas put forward by the project. Also, the project informed—and continues to inform—curriculum design and curriculum documents in mathematics at all levels of the education system in Denmark. An English translation of the most important sections of the KOM report was published in 2011 (Niss and Højgaard 2011).

Concurrently with the KOM project similar ideas emerged elsewhere in the world. To mention just one example, consider the influential *Adding It Up* (National Research Council 2001), produced by the Mathematics Learning Study Committee under the auspices of the National Research Council, edited by Kilpatrick, Swafford and Findell, and published by the National Academies in the USA. In this book we read the following (p. 116):

> Recognizing that no term captures completely all aspects of expertise, competence, knowledge, and facility in mathematics, we have chosen *mathematical proficiency* to capture what we believe is necessary for anyone to learn mathematics successfully. Mathematical proficiency, as we see it, has five components, or *strands*:
>
> - *conceptual understanding*—comprehension of mathematical concepts, operations, and relations
> - *procedural fluency*—skill in carrying out procedures flexibly, accurately, efficiently, and appropriately
> - *strategic competence*—ability to formulate, represent, and solve mathematical problems
> - *adaptive reasoning*—capacity for logical thought, reflection, explanation, and justification
> - *productive disposition*—habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy.
>
> These strands are not independent; they represent different aspects of a complex whole. (National Research Council 2001, p. 116)

Although different in the specifics from the conceptualisation put forward by the competency approach, which focuses on what it takes to *do* mathematics, the approach in *Adding It Up* is an attempt to capture what it takes to *learn* mathematics, and hence what is characteristic of an individual who has succeeded in learning it.

A more recent attempt, in some respects closer to that of the competency approach, can be found in the first part of the US *Common Core State Standards Initiative*, which identifies (2010, pp. 1–2) what is called eight "Standards for Mathematical Practice" common to all (school) levels as below.

- Make sense of problems and persevere in solving them.
- Reason abstractly and quantitatively.
- Construct viable arguments and critique the reasoning of others.
- Model with mathematics.
- Use appropriate tools strategically.
- Attend to precision.
- Look for and make use of structure.
- Look for and express regularity in repeated reasoning.

Since the first inception of the competency approach to mathematics, the KOM project and its ramifications have been subject to a lot of further development and follow-up research in various parts of the world. This, together with experiences gained from various sorts of uses of the competency approach in different places and contexts, has given rise to conceptual and terminological development and refinement. This is not the place to elaborate on these developments. Suffice it to mention that one modification of the scheme is essential in the research done by some of the MEG members to capture and characterise item difficulty in PISA, see the next section and in Chap. 4 of this volume.

Further Aspects of the Notion of Competency

It should be underlined that the eight competencies are not mutually disjoint, nor are they meant to be. (Note differences here with the closely related scheme for item rating in Chap. 4 of this volume.) On the contrary the whole set of competencies has a non-empty intersection. In other words, the competencies do not form a partition of the concept of mathematical competence. Yet each competency has an identity, a 'centre of gravity', which distinguishes it from the other competencies. The fact that all competencies overlap can be interpreted such that the activation of each competency involves a secondary activation of the other competencies, details depending on the context. Consider, for example, the modelling competency. Working to construct a model of some situation in an extra-mathematical context presupposes ideas of what sorts of mathematical questions might be posed in such a context and of what sorts of answers can be expected to these questions. In other words, the thinking competency is activated. Since the very purpose of constructing a mathematical model is to mathematise aspects and traits of the extra-mathematical situation, leading to the posing of mathematical problems that then have to be solved, the problem handling competency enters the scene. Carrying out the problem solving processes needed to solve the problems arising from the mathematisation normally requires the use of mathematical representations, as well as manipulating symbolic expressions and invoking some formalism, alongside using mathematical aids and tools, e.g. calculators or computers, including mathematical software. In other words the representation competency, the symbols and formalism competency, and the aids and tools competency are all activated as part of the process of solving the problem(s) posed. In order to validate, and eventually justify, the solutions and answers obtained as a result of the modelling steps just mentioned, the reasoning competency has to be activated. Finally, beginning and undertaking the modelling task usually requires activation of the receptive side of the communication competency, whereas presenting the modelling process, the model constructed, the model results and their justification, to others activates the constructive side of the communication competency.

In the KOM project we chose to represent the set of competencies as the *competency flower* shown in Fig. 2.1. Each petal represents a competency. They are all distinct petals although they overlap. The density of the shading of each petal is maximal in the middle, at the 'centre of gravity', and fades away towards the boundary. The centre of the flower is the non-empty intersection of all the competencies. Even though a given petal may seem to have a larger intersection with its two neighbours than with the other petals, this is not meant to point to a closer relationship amongst neighbouring petals than amongst other sets of petals.

Possessing a mathematical competency is clearly not an issue of all or nothing. Rather we are faced with a continuum. How, more specifically, can we then describe the extent of an individual's possession of a given competency? The approach taken by the KOM project was to identify three dimensions of the possession of any competency, called *degree of coverage*; *radius of action*; and *technical level*.

A more detailed description of each of the competencies includes a number of aspects employed to characterise that competency. Take, for instance, the representation competency. One of its aspects is to interpret mathematical representations. Another aspect is to bring representations to use, a third is to translate between representations, whereas a fourth aspect is to be aware of the kind and amount of information about mathematical entities that is contained—or left out—in a given representation. Moreover, all of these aspects pertain to any specific mathematical representation under consideration. The *degree of coverage* of a given competency, in this case the representation competency, then refers to the extent to which a person's possession of the competency covers all the *aspects involved in the definition and delineation of that competency*. The more aspects of the competency the person possesses, the higher the degree of coverage of that competency with that person.

Each competency is meant to deal with and play out in challenging mathematics-laden situations that call for the activation of that particular competency. Of course, there is a wide variety of such situations, some more complex and demanding than others. For example, the communication competency can be put to use in situations requiring a person to show and explain how he or she solved a certain task, but it can also be put to use in situations where the person is requested to present and defend his or her view of mathematics as a subject. The *radius of action* of a given competency refers to the *range of different kinds of contexts and situations* in which a person can successfully activate the competency. The wider the variety of contexts and situations in which the person can activate the competency, the larger the radius of action of that competency with that person.

Different mathematics-laden situations give rise to different levels of mathematical demands on a given competency. The symbols and formalism competency, for instance, can be activated in situations that require dealing with arithmetic operations on concrete rational numbers using the rules that govern the operations. It can also be activated, however, in situations that require finding the roots of third degree polynomials, or the solution of separable first order differential equations. The *technical level* on which an individual possesses a given competency, in this case the symbols and formalism competency, refers to the *range of conceptual and technical mathematical demands* that person can handle when activating the competency at issue. The broader the set of demands the person can handle with respect to the competency, the higher the technical level on which the person possesses that competency.

The three dimensions of the possession of a competency allow us to characterise progression in competency possession by an individual as well as by groups or populations. A person's possession of a given competency *increases* from one point in time to a later point in time, if there is an increase in at least one of the three dimensions, degree of coverage, radius of action or technical level, and no decrease in any of them at the same time. This can be extended to groups or entire populations if some notion of average is introduced. Taking stock of the change of average competency possession for all eight competencies across groups or populations allows us to capture progression (or regression for that matter) in mathematical competence at large for those groups or populations. The three dimensions can also be used to compare the intended or achieved mathematical competency profiles of different segments of the education system, or even of different such systems. It is worth noting that such comparisons over time within one section of the education system, or at the same time between segments or systems, attribute at most a secondary role to mathematical content.

One issue remains to be considered. What is the relationship between mathematical competencies and mathematical content? In the same way as it is true that linguistic competencies are neither developed nor activated in environments without the presence of spoken or written language, mathematical competencies are neither developed nor activated without mathematical content. Since one and the same set of mathematical competencies are relevant from kindergarten to university, and vis-à-vis any kind of mathematical content, we can neither derive the competencies from the content, nor the content from the competencies.

The position adopted in the KOM project is that the eight competencies and any set of mathematical content areas, topics, should be perceived as constituting two independent, orthogonal spaces.

Analysis and Discussion of the Role of Competencies Within PISA

It should be borne in mind when reading this section that for all official PISA documents published by the OECD the final authorship and the corresponding responsibility for the text lie with the OECD, even though the international contractors under the leadership of the Australian Council for Educational Research, in turn seeking advice from the Mathematics Expert Group, was always, of course, a major contributor to the publications ‘behind the curtains’.

In the first PISA survey administration, in 2000, mathematics was a minor assessment domain (reading being the major domain). The initial published version of the Framework (OECD 1999), gives emphasis to a version of the eight mathematical competencies of the KOM project. In the text they actually appear as ‘skills’ (‘mathematical thinking skill’, ‘mathematical argumentation skill’, ‘modelling skill’, ‘problem posing and solving skill’, ‘representation skill’, ‘symbolic, formal and technical skill’, ‘communication skill’, and ‘aids and tools skill’) but under the section headed ‘Mathematical competencies’ (p. 43), the opening paragraph uses the term ‘competency’. This is the first indication of reservations and (later) problems with the OECD concerning the term ‘mathematical competency’. In the Framework, ‘mathematical competencies’ was presented as one of two major aspects (p. 42), the other one being ‘mathematical big ideas’, along with two minor aspects, ‘mathematical curricular strands’ and ‘situations and contexts’. Together these aspects were used as organisers of the mathematics (literacy) domain in PISA 2000. Based on the point of view that the individual competencies play out collectively rather than individually in real mathematical tasks (p. 43), it was not the intention to assess the eight competencies individually. Instead, it was decided to aggregate them (quite strongly) into what were then called ‘competency classes’—Class 1: reproduction, definitions, and computations; Class 2: connections, and integration for problem solving; Class 3: mathematical thinking, generalisation and insight. The Framework emphasises that all the skills are likely to play a role in all competency classes. The degree of aggregation of the competencies into competency classes is very high, so that the competency classes take precedence as an organising idea, while the competencies are recognised to play a component role in all mathematical activity.

Soon after, in a precursor publication to the official report of PISA 2000, (OECD 2000) the terms ‘competencies’ and ‘skills’ of the Framework were replaced with the term ‘mathematical processes’ (p. 50). The headings are unchanged, except that the word ‘skill’ is omitted in each of them. Similarly, the ‘competency classes’, including the very term, were preserved but now referred to as ‘levels of mathematical competency’.

The first results of PISA 2000 were officially reported in 2001 (OECD 2001). As to the competencies, they almost disappeared in that report. The notion of mathematical processes as composed of different kinds of skills was preserved. The competency classes of the 1999 Framework were changed to ‘competency clusters’ simply labelled ‘reproduction’, ‘connections’ and ‘reflection’ (p. 23). Apart from that no traces of the competencies are left in the report, including in Chap. 2 in which the findings concerning mathematical literacy are presented.

Mathematics was the major domain in PISA 2003. In the Framework (OECD 2003), it is interesting to observe that the eight mathematical competencies are back on stage in a slightly modified version. In outlining the main components of the mathematics assessment, the Framework reads:

> The process of mathematics as defined by general mathematical competencies. These include the use of mathematical language, modelling and problem solving skills. Such skills, however, are not separated out in different text [sic, should be test] items, since it is assumed that a range of competencies will be needed to perform any given mathematical task. Rather, questions are organised in terms of ‘competency clusters’ defining the type of thinking skill needed. (OECD 2003, p. 16)

This short text, six lines in the original, succeeds in interweaving process, competencies and skills, whilst letting questions be organised by way of competency clusters that define thinking skills. However, in the chapter devoted to mathematical literacy (Chap. 1), there is a clearer—and much more detailed—account of the competencies and their role in the Framework. Taking its point of departure in mathematisation, focusing on what is called, there, ‘the mathematisation cycle’ (p. 38), (and called the modelling cycle in the PISA 2012 Framework (OECD 2013), see also Chap. 1 of this volume) the role of the competencies is to underpin mathematisation. The Framework reads:

> An individual who is to engage successfully in mathematisation in a variety of situations, extra- and intra-mathematical contexts, and overarching ideas, needs to possess a number of mathematical competencies which, taken together, can be seen as constituting comprehensive mathematical competence. Each of these competencies can be possessed at different levels of mastery. To identify and examine these competencies, OECD/PISA has decided to make use of eight characteristic competencies that rely, in their present form, on the work of Niss (1999) and his Danish colleagues. Similar formulations may be found in the work of many others (as indicated in Neubrand et al. 2001). Some of the terms used, however, have different usage among different authors. (OECD 2003, p. 40)

The Framework moves on to list the competencies and their definition. These are ‘Thinking and reasoning’, ‘Argumentation’, ‘Communication’, ‘Modelling’, ‘Problem posing and solving’, ‘Representation’, ‘Using symbolic, formal and technical language and operations’, and ‘Use of aids and tools’ (pp. 40–41). The three competency clusters of the PISA 2000 report (reproduction, connections, and reflection) were preserved in the PISA 2003 Framework, but whilst the competencies didn’t appear in the description of these clusters in PISA 2000, they were indeed present in PISA 2003. For each of the three clusters, the ways in which the competencies manifest themselves at the respective levels are spelled out in the Framework (OECD 2003, pp. 42–44 and 46–47, respectively).

How then, do the competencies figure in the first report on the PISA 2003 results (OECD 2004)? In the summary on p. 26 the competencies as such are absent; only the competency clusters are mentioned. In Chap. 2, reporting in greater detail on the mathematics results, the competencies are only listed by their headings (p. 40) when the report briefly states that they help underpin the key process, identified as mathematisation. In the description of the competency clusters (pp. 40–41) there is no mention of the competencies. Even though competencies are referred to in the previous paragraph (p. 40), they do not appear in the competency clusters. The description of the six levels of general proficiency in mathematics (p. 47) employs some elements from the competency terminology. So, the re-introduction of the competencies into the Framework of PISA 2003 was not really maintained in the reporting of the outcomes.

Apart from what seems to be a general reservation within the OECD towards using the notion of competency in relation to a specific subject—they prefer to use the term to denote more general, overarching processes such as cross-curricular competencies (OECD 2004, p. 29)—there is also a more design-specific and technical reason for the relative absence of the competencies in the report. The classification system for PISA items (that which is called the metadata in Chap. 7 of this volume) did not include information on the role of the eight competencies in the individual items. An item was not classified with respect to all the competencies, only assigned to one of the three competency clusters and other characteristics such as overarching idea, response type etc. This means that there were no grounds on which the PISA results could attribute any role to the individual competencies except in more general narratives such as the proficiency level descriptions. In retrospect one may see this as a deficiency in the Framework. If the eight competencies were to play a prominent role in the design of the PISA mathematics assessment, each of the competencies, and not only the competency clusters, would have to be used in the classification of all the items.

In 2009 the OECD published an in-depth study on aspects of PISA 2003 mathematics done by a group of experts from within and outside the MEG in collaboration with the OECD (2009a). In this report, the eight competencies re-emerge under the same headings as in the 2003 Framework, and with the following opening paragraph:

> An individual who is to make effective use of his or her mathematical knowledge within a variety of contexts needs to possess a number of mathematical competencies. Together, these competencies provide a comprehensive foundation for the proficiency scales described further in this chapter. To identify and examine these competencies, PISA has decided to make use of eight characteristic mathematical competencies that are relevant and meaningful across all education levels. (OECD 2009a, p. 31)

On the following pages (pp. 32–33) of the report, each of the competencies is presented as a key contributor to mathematical literacy.

Science was the major domain in PISA 2006, whereas mathematics was a minor domain so the 2006 Framework (OECD 2006) was pretty close to that of 2003 for mathematics. The central mathematical process was still mathematisation, depicted by way of the mathematisation cycle (p. 95). The competencies were introduced as one of the components in the organisation of the domain:

> The competencies that have to be activated in order to connect the real world, in which the problems are generated, with mathematics, and thus to solve the problems. (p. 79)

Otherwise, the role and presentation of the competencies (pp. 96–98) resembled those of 2003, as did the three competency clusters and the description of their competency underpinnings.

The reporting of the mathematics outcomes of PISA 2006 (OECD 2007) is rather terse, focused on displaying and commenting on a set of items and on presenting the six proficiency levels, the same as used in 2003. In the report, there is no explicit reference to the competencies, even though words from the competency descriptions in the Framework are interspersed in the level descriptions. In this context it is interesting to note that the term 'competencies' does in fact appear in the very title of the report, but in the context of science, "PISA 2006. Science Competencies for Tomorrow's World".

As regards the competencies, the PISA 2009 Mathematics Framework (OECD 2009b) is very close to 2003 and 2006, with insignificant changes of wording here and there. It is interesting, though, that the heading of the section presenting the competencies has been changed to "the cognitive mathematical competencies". The overall reporting of the 2009 mathematics outcomes (OECD 2010) does not deviate from that of 2006. The same is true of the role of the competencies.

In PISA 2012, mathematics was going to be the major domain for the second time. In the course of the previous PISA survey administrations certain quarters around the world had aired some dissatisfaction with the focus on mathematical literacy and with the secondary role attributed to classical content areas in the assessment framework. It was thought, in these quarters, that by assessing mathematical literacy rather than 'just mathematics', the domain became more or less misrepresented. With reference to the need to avoid monopolies, there were also parties in OECD PISA who wanted to diversify the management of PISA, which throughout the life of PISA had taken place in a Consortium (slightly changing over time) led by the Australian Council for Education Research (ACER). Several authors of chapters in this book have personally witnessed expressions of dissatisfaction with aspects of the design of PISA mathematics and an increasing ensuing pressure on those involved in PISA mathematics to accommodate the dissatisfaction.

This is not the place to go into details with evidence and reflections concerning the activities that took place behind the public stage of PISA, but one result of these activities was that PISA mathematics 2012 was launched in a somewhat different setting to what was the case in the previous survey administrations. First, a new agency *Achieve*, from the USA, was brought in to oversee, in collaboration with ACER, the creation of a new Mathematics Framework, especially with regard to the place of mathematical content areas. Secondly, a number of new MEG members were appointed to complement the set of members in the previous MEG which was rather small because mathematics was a minor domain in PISA 2006 and 2009. The opening meeting of the new MEG was attended by a chief officer of the OECD who gave clear indications of the desired change of course with respect to PISA mathematics 2012.

The process to produce a Framework for PISA 2012 mathematics became a lengthy and at times a difficult one, in particular because it took a while for the

MEG to come to a common understanding of the boundary conditions and the degrees of freedom present for the construction of the Framework. After several meetings and iterations of draft texts, the MEG eventually arrived at a common document—submitted to the OECD in the northern autumn of 2010—which was to everyone's satisfaction, even though several compromises had of course to be made, but at a scale that was acceptable to all members, as well as to *Achieve*, ACER and eventually the PISA Governing Board.

Some of the compromises were to do with the competencies and their role in the Framework. We shall take a closer look at these issues below. Before doing so, it is worth mentioning that as the very term 'mathematical competencies' was not acceptable to the OECD for PISA 2012, the term chosen to replace it was 'fundamental mathematical capabilities', whilst it was acknowledged that these had been called 'competencies' in previous Frameworks (OECD 2013, pp. 24 and 30). As will be detailed below, the names, definitions, and roles of these capabilities have, in fact, been changed as well.

Technically speaking the definition of mathematical literacy in the 2012 Framework (p. 25) appeared to be rather different from the ones used in previous Frameworks. However, in the view of the MEG the only difference was that the new definition attempted to explicitly bring in some of the other Framework elements in the description so as to specify more clearly, right at the beginning in the definition, what it means and takes to be mathematically literate. So, the change has taken place on the surface rather than in the substance.

In the introduction to the Framework (OECD 2013, p. 18), the mathematical processes are summarised as follows:

> *Processes*: These are defined in terms of three categories (*formulating situations mathematically*; *employing mathematical concepts*, *facts*, *procedures and reasoning*; *and interpreting*, *apply* [sic] *and evaluating mathematical outcomes*—referred to in abbreviated form as *formulate*, *employ* and *interpret*)—and describe what individuals do to connect the context of a problem with the mathematics and thus solve the problem. These three processes each draw on the seven fundamental mathematical capabilities (*communication*; *mathematising*; *representation*; *reasoning and argument*; *devising strategies for solving problems*; *using symbolic*, *formal and technical language and operations*; *using mathematical tools*) which in turn draw on the problem solver's detailed mathematical knowledge about individual topics.

The role of the fundamental mathematical capabilities—a further modification of the eight mathematical competencies of the KOM project and of the previous four Frameworks—in the 2012 Framework is to underpin the new reporting categories of the three processes (*Formulate—Employ—Interpret*) (see Chap. 1 of this volume.) A detailed account of how this is conceptualised is given on pages 30–31 and in Fig. 1.2 in the Framework (OECD 2013). Apart from the change of terminology from 'mathematical competencies' to 'fundamental mathematical capabilities', which is primarily a surface change, what are the substantive changes involved—signalled by the new headings of the fundamental capabilities—and what caused them? (As 'competency' is the generally accepted term in several quarters outside PISA, we continue to use this term rather than fundamental mathematical capabilities in the remainder of this chapter.) There are three such changes. First, there are some changes in the number and names of the competencies. For example, in the particular context of PISA it was never possible to really disentangle the mathematical thinking competency from the reasoning competency, especially as the former was mainly present indirectly and then closely related to the latter. It was therefore decided to merge them under the heading ‘reasoning and argument’. This change is predominantly of a pragmatic nature.

The second, and most significant, change is in the definition and delineation of the fundamental capabilities. In the first place, this change is the result of research done over almost a decade by members of the MEG with the purpose of capturing and characterising the intrinsic mathematical competency demands of PISA items (see Chap. 4 in this book). The idea is to attach a competency vector, the seven components of which are picked from the integers 0,1,2,3, to each item. Over the years, in this research, it became increasingly important to reduce or remove overlap across the competency descriptions, primarily in order to produce clear enough descriptions for experts to be able to rate the items in a consistent and reliable manner. It was also because the scheme was used to predict empirical item difficulty, which imposed certain requirements in order for it to be psychometrically reliable. This means that the fundamental mathematical capabilities are defined and described in such a way that overlap between them is minimal. This is in stark contrast to the original system of competencies, all of which, by design, overlap. Even though there is a clear relationship between the eight competencies and the seven fundamental mathematical capabilities (e.g. ‘communication’ corresponds to ‘communication’, ‘modelling’ corresponds to ‘mathematising’, ‘thinking and reasoning’ together with ‘argumentation’ correspond to ‘reasoning and argumentation’) the correspondence between the two sets is certainly not one-to-one. In the final formulation of the 2012 Framework it was decided to use the descriptions and delineations from the PISA research project to define the fundamental mathematical capabilities. This implies that the set of mathematical competencies does not make the set of fundamental mathematical capabilities superfluous, nor vice versa. They have different characteristics and serve different purposes, namely providing a general notion of mathematical competence and a scheme to analyse the demands of PISA items, respectively. From that perspective it can be seen as a stroke of luck that the requirement to introduce a new terminology eventually served to avoid confusion of the scheme of the KOM project (and the earlier versions of the PISA Framework) and the 2012 Framework.

The third change was one of order. The fundamental mathematical capabilities of the 2012 Framework occur in a different order than did the mathematical competencies of the previous survey administrations. The reason for this reordering was an attempt to partially (but not completely) emulate the logical order in which a successful problem solver meets and approaches a PISA item. First, the problem solver reads the stimulus and familiarises himself or herself with what the task is all about. This requires the receptive part of ‘communication’. Next, the problem solver engages in the process of mathematising the situation (i.e. ‘mathematising’), whilst typically making use of mathematical representations (i.e. ‘representation’) to come to grips with the situation, its objects and phenomena. Once the situation has been mathematised, the problem solver has to devise a strategy to solve the ensuing mathematical problems (i.e. ‘devising strategies for solving problems’). Such a strategy will, more often than not, involve ‘using symbolic, formal and technical language and operations’, perhaps assisted by ‘using mathematical tools’. Then comes an attempt to justify the solutions and mathematical conclusions obtained by adopting the strategy and activating the other capabilities (i.e. ‘reasoning and argument’). Finally the problem solver will have to communicate the solution process and its outcome as well as its justification to others. This takes us back to ‘communication’, now to its expressive side.

At the time of writing this chapter, the official report of PISA 2012 had not yet been published. So, it is not possible to consider the way in which the three processes and the fundamental mathematical capabilities fare in the reporting. This is, of course, even more true of PISA 2015 and subsequent PISA survey administrations, which are in the hands of a completely different management, even though my role as a consultant to the agency in charge of producing the PISA 2015 Framework allows me to say that this Framework is only marginally different from the PISA 2012 Framework.

Concluding Remarks

This chapter has attempted to present the genesis, notion and use of mathematical competencies in Denmark and in other places with a side view to analogous ideas and notions, so as to pave the way for a study of the place and role of mathematical competencies and some of their close relatives, fundamental mathematical capabilities, in the Frameworks and reports of the five PISA survey administrations that at the time of writing have almost been completed (September 2013). The chapter will be concluded by some remarks and reflections concerning a special but significant issue of the relationship between competencies (capabilities) and the entire Framework. In a condensed form this issue can be phrased as a question: ‘what underpins what?’

From the very beginning of PISA the approach to the key constituent of the mathematics assessment, i.e. mathematical literacy, was based on mathematical modelling and mathematisation of situations in contexts, although the specific articulation of this in the Framework varied from one survey administration to the next, as did the related terminology. In other words, modelling and mathematisation were always at the centre of PISA. However, the eight mathematical competencies, and most recently the seven fundamental capabilities, were part of the Frameworks as well. Now, do we detect here a potential paradox or some kind of circularity, since modelling (mathematising) is one of the eight competencies (seven capabilities) underpinning the whole approach, above all modelling? It is not exactly surprising that a set of competencies that includes modelling can serve to underpin modelling. If modelling is in centre, why do we need the other competencies then?

Alternatively, would it have been better (if possible) to specify mathematical literacy in terms of the possession of all the mathematical competencies, without focusing especially on the modelling competency, the possession of which would then become a corollary?

Let us consider the first question. When it comes to the eight competencies, it was mentioned in a previous section that the fact that they all overlap means that even when the emphasis is on one of the competences, the others enter the field as 'auxiliary troops' in order for the competency at issue to be unfolded and come to fruition. It is therefore consistent with this interpretation to have the entire system of competencies underpin the modelling competency. One might say, though, that were it only for PISA, in which the emphasis is on the modelling competency, that competency might have been omitted from the list in order to avoid the tiny bit of circularity that is, admittedly, present. However, as the competency scheme is a general one used in a wide variety of contexts, and not only in PISA, it would be unreasonable to remove it from the list solely because of its special use in PISA. What about the seven fundamental mathematical capabilities in the 2012 Framework, then? Here the circularity problem has actually disappeared, at least terminologically speaking, because the seven capabilities do not contain one called modelling, only mathematising (and in a more limited sense than it sometimes has), and because the term mathematising is not used in the modelling cycle in the Framework, as it has been replaced by 'formulating situations mathematically'. So, in the 2012 Framework it is indeed the case that the capabilities underpin this process as well as the other two, 'employing mathematical concepts, facts procedures and reasoning', and 'interpreting, applying and evaluating mathematical outcomes'.

As to the second question, since the eight competencies are meant to constitute mathematical competence and mastery at large, the option mentioned would have amounted to equating mathematical literacy and mathematical competence. This is certainly a possible but not really a desirable option. The perspective adopted in PISA, right from the outset, was not to focus on young people's acquisition of a given subject, such as mathematics, but on their ability to navigate successfully as individuals and citizens in a multifaceted society as a result of their compulsory schooling. This zooms in on putting mathematics to use in a variety of mainly extra-mathematical situations and contexts, in other words the functional aspects of having learnt mathematics. This is what mathematical literacy is all about, being brought about by way of modelling. I, for one, perceive mathematical literacy as a proper subset of mathematical competence, which implies that for someone to be mathematically competent he or she must also be mathematically literate. Even though mathematical literacy does indeed draw upon (aspects of) all the competencies, it does not follow that all the competencies are represented at a full scale and in an exhaustive manner. So, the converse implication, that a mathematically literate person is also necessarily mathematically competent, does not hold. Mathematical competence involves operating within purely mathematical structures, studying intra-mathematical phenomena such as the irrationality of $\sqrt{2}$ and π

even though this is never really required in the physical world, and at a higher level understanding the role of axioms, definitions and proofs.

These remarks are meant to show that what at face value may appear, to some, as a kind of circularity or inconsistency in the PISA Frameworks concerning mathematical literacy, mathematical competence and competencies, fundamental mathematical capabilities, modelling and mathematising are, as a matter of fact, basically logically coherent in a closer analysis.

It will be interesting to follow, in the years to come, how mathematical competencies are going to be developed from research as well as from practice perspectives. At the very least, putting the competencies on the agenda of mathematics education has offered new ways of thinking about what mathematics education is all about.

References

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. www.corestandards.org/Math. Accessed 30 Aug 2013.

Devlin, K. (1994). *Mathematics, the science of patterns*. New York: Scientific American Library.

Devlin, K. (2000). The four faces of mathematics. In M. J. Bourke & F. R. Curcio (Eds.), *Learning mathematics for a new century. The NCTM 2000 yearbook* (pp. 16–27). Reston: National Council of Teachers of Mathematics.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education. China lectures*. Dordrecht: Kluwer.

Husén, T. (Ed.). (1967). *International study of achievement in mathematics, a comparison of twelve countries* (Vols. I & II). New York: Wiley.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston: The National Council of Teachers of Mathematics.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.), Mathematics Learning Study Committee. Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Niss, M. (1999). Kompetencer og uddannelsesbeskrivelse [Competencies and subject descriptions]. *Uddannelse* [Education], *9*, 21–29

Niss, M. (2001). University mathematics based on problem-oriented student projects: 25 years of experiences with the Roskilde Model. In D. Holton (Ed.), *The teaching of learning of mathematics at university level. An ICMI study* (pp. 153–165). Dordrecht: Kluwer.

Niss, M., & Højgaard, T. (Eds.). (2011). *Competencies and mathematical learning. Ideas and inspiration for the development of mathematics teaching and learning in Denmark* (Tekster fra IMFUFA, no 485). Roskilde: Roskilde University, IMFUFA.

Niss, M., & Jensen, T. H. (Eds.). (2002). *Kompetencer og matematiklæring. Ideer og inspiration til udvikling af matematikundervisning i Danmark*. Uddannelsesstyrelsens temahæfteserie nr. 18. Copenhagen: Ministry of Education.

Niss, M., Emanuelsson, J., & Nyström, P. (2013). Methods for studying mathematics teaching and learning internationally. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 975–1008). New York: Springer.

Organisation for Economic Co-operation and Development (OECD). (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2000). *Measuring student knowledge and skills. The PISA 2000 assessment of reading, mathematical and scientific literacy. Education and skills*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2001). *Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2003). *The PISA 2003 assessment framework—Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2004). *Learning for tomorrow's world—First results from PISA 2003*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2007). *Science competencies for tomorrow's world* (Analysis, Vol. 1). Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2009a). *Learning mathematics for life: A perspective from PISA*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2009b). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics and science* (Vol. 1). Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2013). *PISA 2012 assessment and analytical framework. Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didakik der Mathematik, 61*, 37–46.

Chapter 3
The Real World and the Mathematical World

Kaye Stacey

Abstract This chapter describes the way in which PISA theorises and operationalises the links between the real world and the mathematical world that are essential to mathematical literacy. Mathematical modelling is described and illustrated and the chapter shows why it is used as the cornerstone to mathematical literacy. It discusses how this concept has developed over the PISA Frameworks from 2000 to 2012, culminating in the reporting in PISA 2012 of student proficiency in the three modelling processes of *Formulate*, *Employ* and *Interpret*. Consistent with the orientation to mathematical modelling and mathematisation, the authenticity of PISA items is given a high priority, so that students feel that they are solving worthwhile, sensible problems. The use of real-world contexts is regarded as essential to teaching and assessing mathematics for functional purposes and in assisting in motivation of students, but potential problems of cultural appropriateness and equity (through familiarity, relevance and interest) arise for an international assessment. This is the case for countries as a whole and also for subgroups of students. Relevant research and the PISA approach to minimising potential biases are discussed.

Introduction

The emphasis of PISA's mathematical literacy is on "mathematical knowledge put to functional use in a multitude of different situations" (OECD 2004, p. 25). It follows from this that presenting students with problems in real-world contexts is essential. PISA has steered away from the dubious route of inferring students' ability to solve problems in real-world contexts from a measure of students' ability to perform mathematical procedures in the abstract (e.g. solving equations, performing calculations). The use of real-world contexts and how this interacts with the world of mathematics is therefore the theme of this chapter.

K. Stacey (✉)
Melbourne Graduate School of Education, The University of Melbourne,
234 Queensberry Street, Melbourne, VIC 3010, Australia
e-mail: k.stacey@unimelb.edu.au

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_3

Within the mathematics education world, the process of applied problem solving (solving problems that are motivated by a concern arising outside of the world of mathematics itself) has for many years been widely described by means of the 'mathematical modelling cycle' (Blum and Niss 1991). The process of mathematical modelling is described in this chapter, which discusses the concept from the theoretical perspective as well as explaining in detail how it is linked to mathematical literacy and PISA items.

Whereas mathematising the real world and using mathematical modelling to solve problems always been a cornerstone of PISA (although variously named in the various surveys), this was not evident in the reporting of PISA results, which gave only overall scores for mathematical literacy and scores for the four content categories (*Space and shape*; *Quantity* etc.). However, in PISA 2012 the modelling cycle has also been used to provide an additional reporting category for student proficiency. The major reason for this was to describe more precisely what proficiencies make up mathematical literacy, and to report how well different groups of students do on each of these. More detailed reporting gives educational jurisdictions better information from PISA about the strengths of their students.

The PISA 2009 survey of science (OECD 2010) reported the degree to which three scientific competencies are developed: identifying scientific issues, explaining phenomena scientifically and using scientific evidence. What is a parallel way of thinking about the constituents of mathematical literacy? The answer, from the modelling cycle, is discussed in this chapter. The purpose of this chapter is therefore:

- to describe mathematical modelling and to show why it is the key to PISA's mathematical literacy
- to demonstrate with sample PISA items how mathematical literacy is connected with modelling
- to discuss the reporting in PISA 2012 according to the three mathematical processes of *Formulate*, *Employ* and *Interpret*
- to link mathematical literacy and mathematical modelling with mathematisation
- to discuss item design issues concerning the use of real-world contexts in PISA problems, especially related to authenticity and equity.

Mathematics is a difficult subject to learn because all mathematical objects are abstract: numbers, functions, matrices, transformations, triangles. Even though we can identify triangle-like shapes around us, we cannot see the abstract 'object' of a triangle; we must impose the mental concept of triangle on the real-world thing. Perhaps surprisingly mathematics derives much of its real-world power from being abstract: abstract tools developed in one context can be applied to many other physical phenomena and social constructs of the worlds of human experience and science. This is what mathematical modelling does. A problem arising in the 'real world' is transformed to an intra-mathematical problem that can be solved (we hope) using the rules that apply to abstract mathematical objects and which may have been first derived or discovered for quite a different area of application. Then the solution is used for the real-world purposes. This real world includes personal, occupational, societal and scientific situations, not just physical situations; a convention that is summarised in PISA's *Personal*, *Societal*, *Occupational* and *Scientific* context categories (see Chap. 1). Critically also and perhaps paradoxically, the real world for mathematics is not confined to what actually exists. Ormell (1972) describes the greatest value of mathematics as providing, through its modelling capability, the ability to look at possibilities; testing out the details of not-yet-actualised situations. A great deal of investigation of the feasibility and necessary characteristics of the sails described in PM923 Sailing ships (see Chap. 1 of this volume) would be done mathematically, long before any sail is manufactured.

Mathematical Modelling

What Is a Mathematical Model?

In the past, a mathematical model was a physical object, often something beautiful to be admired or used for teaching. For example, Cundy and Rollett's book entitled "Mathematical Models" (1954) gave detailed instructions for making a wide variety of mathematical models, such as Archimedean and stellated polyhedra and linkages, and for drawing loci. Now, reflecting common usage, the Wikipedia article on mathematical models briefly dismisses this former understanding in one sentence. Instead the article defines a mathematical model as "a description of a system using mathematical concepts and language" and explains the purposes of modelling as "A model may help to explain a system and to study the effects of different components, and to make predictions about behaviour." One quick search of an online job advertisement agency using the term "mathematical modelling" showed that there are vacancies today in my region for mathematical modellers in banking, finance and accounting, agriculture, gambling and online gaming, mechanical engineering, software engineering, marketing, mining and logistics. It is clear from this that mathematical modelling is essential to business and industry.

The primary meaning of the word 'model' (as a noun) now refers to

- the set of simplifying assumptions (e.g. which variables are important in the situation for the problem at hand, what shape something is),
- the set of assumed relationships between the variables, and
- the resulting formula or computer program or other device that is used to generate an answer to the question.

Models can be extraordinarily complex, such as the highly sophisticated models that are used for predicting the weather. They can summarise profound insights into the nature of the universe, such as Newton's three laws of motion. Models can also be very simple, like many of the rules of thumb and instructions that we use on a daily basis. I make tea in a teapot by remembering the rule "one [spoonful of tea] for

Conventional Oven. Put sausage rolls on tray in centre of oven. Heat approximately 25 minutes or until hot right through. To heat when unfrozen, reduce heating time to 15 minutes.

Microwave Oven. Microwave on full power for required time. 2 sausage rolls for 1½ minutes. 4 sausage rolls for 2½ minutes. 6 sausage rolls for 3½ minutes. Allow to stand for one minute. Serve.

Fig. 3.1 Instructions for heating on a packet of Aunty Betty's frozen sausage rolls

each person and one for the pot". This is a simple linear model taught to me by my grandmother, based on assumptions of the volume of teapots and preferred strength of tea, and validated by experience. I drive keeping a gap of 2 seconds between the next car and mine: an easily memorised and implemented rule to follow (especially as it is independent of speed) that has been derived from the relationship between stopping distances and speed and based on assumptions about good driving conditions, typical braking force, reaction time etc. Figure 3.1 shows the instructions written on a packet of frozen sausage rolls. For the microwave oven, the time is modelled as a linear function of the number of sausage rolls. For a conventional oven, the model for the heating time is independent of this variable. These mathematically distinct models reflect the very different physical processes of heating in the two ovens, by exciting water molecules with microwaves or from a heat source. They also rely on many simplifying assumptions and relationships, including the size, shape and ingredients of the sausage rolls, the heating capacity of ovens, and food safety (hot enough on the inside to kill germs, but not too hot to burn the mouth). Of course, Aunty Betty herself, in designing the instructions, probably adopted an empirical method, heating sausage rolls and testing the temperature against food safety rules (also perhaps expressed as mathematical models). The normal consumer just needs to follow the instructions to work out the cooking time; a caterer may need to modify the rule for heating a very large number of sausage rolls. Many of the real situations in which mathematical literacy is required arise in the role of 'end user' of a model.

The Modelling Cycle and PISA's Model of Modelling

M154 Pizzas was released after the PISA 2003 survey (OECD 2006b, 2009b). It illustrates the main features of mathematical modelling in a simple way. For anyone feeding a large group of hungry people with pizza, this is a real-world problem. In my city, pizza diameters are often advertised alongside the cost. Note that a zed is the unit of currency in the imaginary Zedland where PISA items are often set, in order to standardise the numerical challenges for students around the world.

M154 Pizzas. A pizzeria serves two round pizzas of the same thickness in different sizes. The smaller one has a diameter of 30cm and costs 30 zeds. The larger one has a diameter of 40cm and costs 40 zeds.

M154Q01. Which pizza is better value for money? Show your reasoning.

A solution involves taking the real-world concept of value for money and describing it mathematically: perhaps as area of pizza per zed (or alternatively zeds per square centimetre, volume per zed, zeds per cubic centimetre). Assuming that the pizza is circular completes the formulation stage: the real-world problem has been transformed into a mathematical problem. Next the calculations can proceed (exactly or approximately) and the comparison of areas of pizza per zed (say) can be made. This is the stage where mathematical techniques come to the fore, in solving the mathematical problem to obtain a mathematical result. After this, the desired real-world solution is identified (the pizza with higher numerical area per zed) and interpreted as a decision that the larger size is better value for money. (Of course, the problem can also be solved algebraically without any calculation comparing the quadratic growth of area with diameter with the linear growth of cost, and similar modelling considerations apply). Next the real-world adequacy and appropriateness of the solution is examined. If only large pizzas are purchased, can everyone get the menu choice that they want? Will too much be purchased? This means that the idea of value for money may need to be more complex than square centimetres per zed. Where M154 Pizzas stops, in real life a new modelling cycle may begin with modified variables, assumptions and relationships (e.g. at least five different pizzas are required for this party) to better aid the “well-founded judgments and decisions” that feature in PISA’s definition of mathematical literacy (OECD 2013a).

When mathematics was first a major domain for the PISA survey in 2003, the Framework (OECD 2004) included a model of the modelling cycle (although there it was called the mathematisation cycle following the RME tradition as in de Lange (1987)). This cycle described the stages through which solving a real-world problem proceeds. Figure 3.2 shows the graphics depicting it that appeared in the 2006

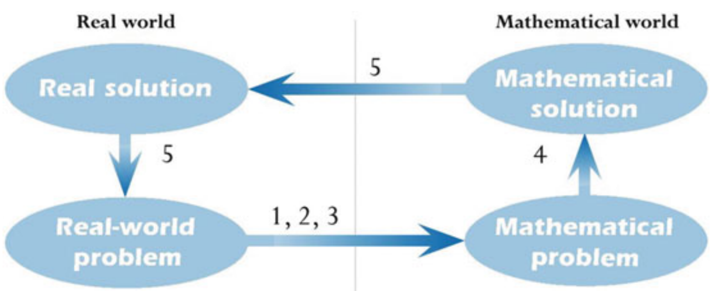


Fig. 3.2 The mathematisation (modelling) cycle (OECD 2006a, p. 95)

Framework (OECD 2006a). Models of the modelling (mathematisation) cycle have long been used in discussing its teaching and learning and there are many variations, which bring in various levels of detail (e.g. Blum et al. 2007; Stillman et al. 2007). A diagram that depicts the modelling cycle in essentially the same way as PISA does was published by Burkhardt in 1981 and there may be earlier occurrences.

The first feature of this diagram is the division into two sides. On the real world side, the discourse and thinking are concerned with the concrete issues of the context (pizzas, money). On the mathematical world side, the objects are abstract (area, numbers) analysed in strictly mathematical terms. Within the ovals are the states that the modelling cycle has reached, and the arrows indicate the processes of movement between these states. The numbers on the diagram give an explanation of the activities that constitute the arrows. The first arrow (labelled (1), (2), (3)) represents the formulation process during which the mathematical features of the situation are identified and the real-world problem is transformed into a mathematical problem that faithfully represents the situation: (1) starting with a problem situated in reality, (2) organising it according to mathematical concepts and identifying the relevant mathematics involved and (3) trimming away the reality by making assumptions, generalising and formalising. The problem solver has thus moved from real-world discourse to mathematical-world discourse. The 'problem in context' (best value for money) has been transformed into a mathematical problem about abstract mathematical objects (area, numbers, rates) that is hopefully amenable to mathematical treatment. The arrow within the mathematical world (4) represents solving the mathematical problem (calculating then comparing the areas per zed). The arrow labelled (5) indicates the activity of making sense of the mathematical solution in terms of the real situation, and considering whether it answers the real problem in a satisfactory way (e.g. large pizzas may not give enough variety).

A more picturesque description of the same modelling cycle was given by Synge:

> The use of applied mathematics in its relation to a physical problem involves three steps. First, a dive from the world of reality into the world of mathematics; two, a swim in the world of mathematics; three, a climb from the world of mathematics back into the world of reality, carrying the prediction in our teeth. (Synge 1951, p. 98)

Apart from the diagram having undergone reflection in a horizontal axis, Fig. 3.2 is extremely similar to Fig. 3.3, which shows the diagram and terminology for the modelling cycle used in the PISA 2012 Framework. In labelling the arrows, the PISA 2012 diagram links directly to the reporting of student proficiency in the separate processes that will be discussed below. There are two arrows that move between the real world and the mathematical world: *Formulate* and *Interpret*. The *Employ* arrow represents solving actions that lie entirely within the mathematical world. Within the real world is the *Evaluate* arrow. Here the result obtained from the model is judged for its adequacy in answering the real-world problem.

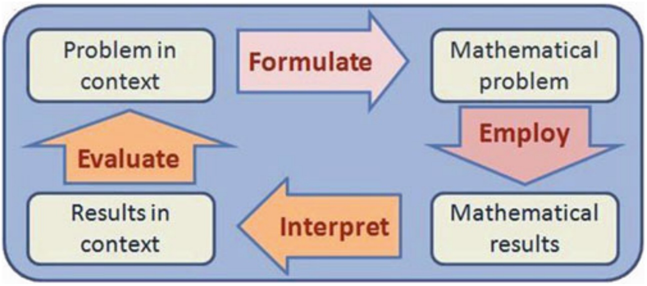


Fig. 3.3 PISA 2012 model of mathematical modelling (OECD 2013a)

If the solution is satisfactory, the modelling ends. If it needs improvement, a modified problem in context has been established, and the cycle begins again probably building in assumptions and relationships that better reflect the real situation.

Both Figs. 3.2 and 3.3 depict an idealised and simplified model of solving a real-world problem with mathematics. In reality, problem solvers can make many movements back and forth rather than steadily progressing forward through the modelling cycle. A result may be found to be unrealistic at the evaluation stage leading to a move forward to a new formulation or instead there may be a move backwards to check calculations or carry them out with greater precision. A formulated model may lead to equations that cannot be solved, prompting a move backwards from *Employ* to *Formulate* to search for assumptions and relationships that will lead to a more tractable mathematical problem. Indeed, the *Formulate* and *Employ* processes need to be closely intertwined because in formulating a mathematical model the problem solver is wise to keep an eye on the technical challenges that lie ahead.

In addition to these back and forth movements between processes, there are deeper ways in which the simple division into the real world and the mathematical world does not reflect reality. Reasoning from the context can be an aid to finding the mathematical solution ("I must have made a mistake because I know mass does not affect the result, so the *m*'s in my formula should cancel"). Furthermore, understanding details of the mathematical solution can be essential to interpreting the findings sensibly (e.g. "I ignored the quadratic terms so I could solve the equations, so it is not surprising that my results show that the quantities are linearly related."; "I assumed cars go through the traffic lights at a rate of 30 per minute, so it is not surprising that as the time that the lights are set on green increases, the number of cars that could pass through the lights tends asymptotically towards 30×60 per hour".)

The mathematical modelling cycle is also affected when people work together, perhaps in employment, with some people creating models and others using them possibly in a routine way. Not all use of mathematics involves the full modelling cycle, which is the key observation when discussing the link between mathematical literacy and modelling below.

Mathematical Literacy and Mathematical Modelling

What is the relationship between PISA's mathematical literacy and mathematical modelling, which is described as its cornerstone and key feature (OECD 2013a)? Two facts are immediately clear. On the one hand, almost by definition, mathematical modelling and mathematical literacy are strongly connected. The definition of mathematical literacy (OECD 2013a) includes to "describe, explain and predict phenomena" and to assist in making "well-founded judgements and decisions". The Wikipedia modelling page quoted above includes a very similar list: "explain a system, study the effects of different components, and to make predictions about behaviour." On the other hand, most people in real life, and especially 15-year-old students working under test conditions, would only rarely engage in the full modelling cycle as described above except in very simple instances of it. For example, only mathematically adept customers probably consider the functional variation described above when buying pizzas, and then probably only if they have to buy a lot. It is, however, much more critical that the pizzeria owner understands the mathematical model for ordering ingredients and setting prices. What is the resolution to this paradox that mathematical modelling is key to mathematical literacy, that everyone needs mathematical literacy, yet most people rarely engage in the whole modelling cycle? In most cases, people exercising their mathematical literacy are engaged in just a part of the modelling cycle with other parts greatly abbreviated. Examples follow.

In very many instances where mathematical literacy is required, people use mathematical models that are supplied to them, greatly truncating the *Formulate* process. Using the 'rule-of-thumb' models referred to above are simple examples. I want to heat five sausage rolls in the microwave. I read the instructions on the packet. Implicitly I assume linear interpolation, so I just have to calculate the time halfway between the times for four and six sausage rolls. Some PISA items are of this 'using models' type. An *Occupational* example, the item PM903Q03 Drip rate Question 3 (OECD 2013a) requires calculation of the volume of a drug infusion given the drip rate, the total time, the drop factor and a formula that connects these four variables together. In a question such as this, the *Formulate* and *Interpret* processes are greatly truncated and the cognitive demand comes almost entirely from the *Employ* process (substituting values, changing the subject of the formula, and calculation).

In many other instances where mathematical literacy is required, the formulation process is greatly truncated because the relevant mathematical models have been explicitly taught and practised at school (e.g. calculating distance from speed and time, area of composite shapes, converting units, percentage discounts for shopping, using scales on maps, reading a pie chart). A very common instance in PISA, as in real life, is where proportional reasoning is required. M413Q01 Exchange Rate Question 1 (OECD 2006a, 2009a) stated that 1 Singapore dollar was worth 4.2 South African rand and asked how many South African rand would be received for 3,000 Singapore dollars. It was the third easiest item in the PISA 2003 survey

(OECD 2009a). The cognitive demand for formulating this problem is very low because conversion of units is a commonly taught application of rate (proportional reasoning), and because the item is set up to go directly from 1 SGD to 3000 SGD.

Reading information from charts and graphs is a common instance of mathematical literacy for citizens and employees, and there are many PISA items testing this, such as PM918Q02 Charts Question 2 (see Fig. 3.4). Items like this almost exclusively involve the *Interpret* process of the modelling cycle. (Note that the interpret process does not involve the receptive communication of reading the question, but is about understanding the real-world meaning of the results.) Relevant mathematical information is presented (often in a graph, a timetable, a diagram) and has to be used quite directly with little processing to answer a question of interest. PM918Q02 Charts Question 2 was an easy item with 79 % of students correct in the field trial. To link this into the modelling cycle, I imagine that this information has been assembled, perhaps by a newspaper or by a sales team. They have formulated the situation mathematically by making a series of choices (e.g. what and how many variables, aggregation by month better than by week, selecting a clustered column graph) and then creating a graph. The end user (perhaps a band manager) and in this case also the PISA test taker exhibiting mathematical literacy, has to interpret this mathematical product, selecting the two data series in question, and compare the heights of the columns visually, starting from January. This activity lies just at the end point of the modelling cycle. In summary, using mathematical literacy can involve full engagement with the mathematical modelling cycle, but most frequently it involves just a small part of it in real life and in PISA.

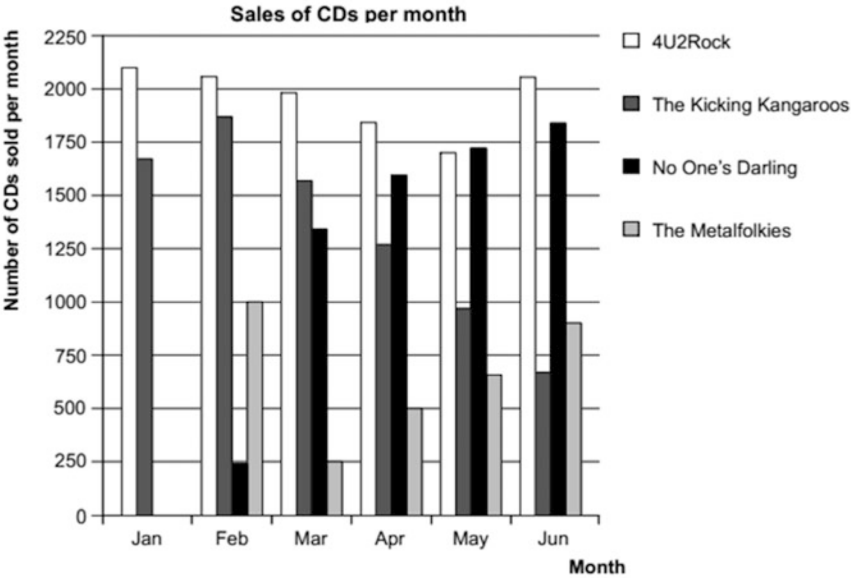
PISA Assessment and the Modelling Cycle

As noted above, in PISA 2012 the modelling cycle has been used to provide a new reporting category. The intention is to describe what abilities make up mathematical literacy and the degree to which students possess them. As discussed in Chap. 2, this is well described by the fundamental mathematical capabilities (called competencies in Chap. 2 and earlier Frameworks), and Turner, Blum and Niss in Chap. 4 provide empirical evidence for this claim. However, reporting against six or more capabilities is impractical because there are just too many and also because they normally occur together in problem solving.

Instead, PISA 2012 uses the processes *Formulate—Employ—Interpret* of the modelling cycle for reporting. All three can generally be identified in solving a problem, but because of the constraints of the PISA assessment (e.g. time) it is nearly always possible to identify that the main demand of an item lies with one of them. As noted in the section above, this also reflects much use of mathematics in real life: some aspects of the modelling cycle are so truncated as to be barely present for the end user. Items that mainly focus on the arrow labelled *Formulate* in Fig. 3.3 are used to measure student performance in *Formulating situations mathematically*.

CHARTS

In January, the new CDs of the bands *4U2Rock* and *The Kicking Kangaroos* were released. In February, the CDs of the bands *No One's Darling* and *The Metalfolkies* followed. The following graph shows the sales of the bands' CDs from January to June.



Question 2: CHARTS *PM918Q02*

In which month did the band *No One's Darling* sell more CDs than the band *The Kicking Kangaroos* for the first time?

A No month
B March
C April
D May

Question 5: CHARTS *PM918Q05*

The manager of *The Kicking Kangaroos* is worried because the number of their CDs that sold decreased from February to June.

What is the estimate of their sales volume for July if the same negative trend continues?

A 70 CDs
B 370 CDs
C 670 CDs
D 1340 CDs

Fig. 3.4 Two questions from the unit PM918 Charts (OECD 2013b)

Items that focus on the *Employ* arrow are used to report on the process formally labelled *Employing mathematical concepts*, *facts*, *procedures*, *and reasoning*. Finally, one process *Interpreting*, *applying*, *and evaluating mathematical outcomes*, abbreviated to *Interpret*, is constructed from items that mainly focus on the interpreting and evaluating arrows. These have been combined because the opportunities for any serious evaluation under the conditions of a PISA survey are severely limited: items are completed in a short time by students sitting at a desk without additional resources.

Above, examples of PISA items that are close to real-world situations and were very strongly focused on just one process were given: PM903Q03 Drip rate Question 3 and M413Q01 Exchange Rate focused on the *Employ* process and PM918Q02 Charts Question 2, focused on the *Interpret* process. M537Q02 Heart beat Question 2 (OECD 2006a, 2009a) is an example of an item strongly focused on the *Formulate* process. The stimulus gave the formula

$$\text{recommended maximum heart rate} = 208 - (0.7 \times age)$$

and the information that physical training is most effective when heartbeat is at 80 % of the recommended maximum. The question asked for a formula for the heart rate for most effective physical training expressed in terms of age. In this item, full credit was obtained by students who left the expression without expansion. For example, both of the equations $heart\ rate = (208 - 0.7 \times age) \times 0.8$ and $h = 166 - 0.6a$ were scored with full credit. Consequently, the main cognitive demand was focused in formulating the new model.

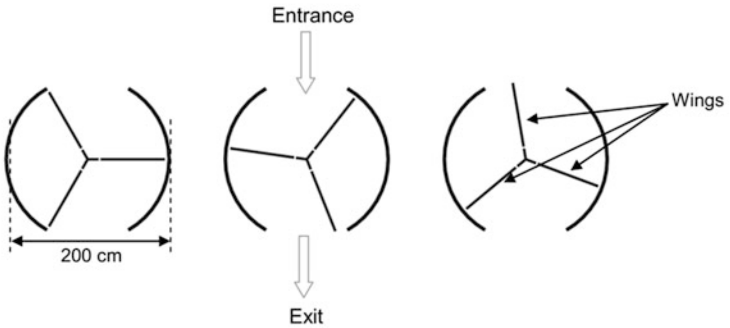
The above PISA items are easy to allocate to just one process, but not all items are like this. The psychometric model used by PISA requires that items be allocated to only one of the three processes, so the following examples illustrate how on-balance judgements are made for items involving more of the modelling process. Three straightforward decisions are illustrated first, followed by the difficult case of PM918Q05 Charts Question 5.

PM995Q03 Revolving Door Question 3 (see Fig. 3.5) involves proportional reasoning, but this item is far from a routine application. Students have to construct a model of the situation (probably implicitly) to go from total time (30 min) to total revolutions (120) to total entry options (360) to total people (720). Although each of these relationships is a standard proportional reasoning situation, they need to be assembled systematically to solve the problem. The item is classified as *Formulate* because the demand from this process was judged to be greater than from the calculation (*Employ*) and interpreting of the answer in real-world terms is very straightforward (*Interpret*).

The item PM995Q02 Revolving door Question 2 (see Fig. 3.5) was one of the most difficult items in the field trial, with only 4 % of students successful. This item makes heavy demands at the formulation stage. It addresses the main purpose of revolving doors, which is to provide an airlock between inside and outside the building and it requires substantial geometric reasoning followed by accurate calculation. The real situation has to be carefully analysed and this analysis needs

REVOLVING DOOR

A revolving door includes three wings which rotate within a circular-shaped space. The inside diameter of this space is 2 metres (200 centimetres). The three door wings divide the space into three equal sectors. The plan below shows the door wings in three different positions viewed from the top.



Question 1: REVOLVING DOOR

PM995Q01 – 0 1 9

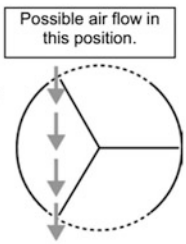
What is the size in degrees of the angle formed by two door wings?

Size of the angle: ....°

Question 2: REVOLVING DOOR

PM995Q02 – 0 1 9

The two door **openings** (the dotted arcs in the diagram) are the same size. If these openings are too wide the revolving wings cannot provide a sealed space and air could then flow freely between the entrance and the exit, causing unwanted heat loss or gain. This is shown in the diagram opposite.



What is the maximum arc length in centimetres (cm) that each door opening can have, so that air never flows freely between the entrance and the exit?

Maximum arc length: .... cm

Question 3: REVOLVING DOOR

PM995Q03

The door makes 4 complete rotations in a minute. There is room for a maximum of two people in each of the three door sectors.

What is the maximum number of people that can enter the building through the door in 30 minutes?

A. 60
B. 180
C. 240
D. 720

Fig. 3.5 The unit PM995 Revolving door (OECD 2013b) to be translated into geometric terms and back again to the contextual situation of the door multiple times during the solution process. As the diagram supplied in the question shows (see Fig. 3.5) air will pass from the outside to the inside, or vice versa, if the wall between the front and back openings is shorter than the circumference subtended by one sector. Since the sectors each subtend one third of the circumference, and there are two walls, together the walls must close at least two thirds of the circumference, leaving no more than one third for the two openings. Assuming symmetry of front and back, each opening cannot be more than one sixth of the circumference. There is further geometric reasoning required to check that the airlock is indeed maintained if this opening length is used. The question therefore draws very heavily on the *reasoning and argument* fundamental capability. It is unclear in this problem when the formulation ends and the employing process begins, because of the depth of geometric reasoning required. A careful analysis of the solution of an individual in terms of the modelling cycle would probably find it often moving from the *Formulate* arrow (what does it mean in mathematical terms to block the air flow?) to the *Employ* arrow and back again. The decision to place this item in the *Formulate* process indicates a judgement that the most demanding aspect is to translate into geometric terms the requirement that no air pass through the door. However, working within the mathematical world is also demanding in this case. Allocating to *Formulate* is supported by the observation that it is more likely that a student will have failed to make progress on this item in the *Formulate* process, rather than have succeeded there and been unable to solve the intra-mathematical problem.

PM918Q05 Charts Question 5 (see Fig. 3.4) illustrates that the allocation to one of the three processes is sometimes unexpectedly complex. To solve this problem, first the phrase “same negative trend” needs to be formulated mathematically, and there are several choices. Formulating graphically might lead the student to physically or mentally draw a line of best fit through the tops of the Kicking Kangaroos columns for February to June, extend the line to where July would be and observe that it will be of height not much below 500 (hence answer B correctly). Alternatively, a gradient for the line could be calculated and applied to calculate a value for July. Formulating numerically, a student may calculate an average drop per month and reduce the June sales by this amount. The interpretation of the answer obtained by any of these processes is simple. The test designers allocated this item to the *Employ* process, deciding that the main cognitive demand is in carrying out any one of these strategies, rather than in deciding that the drop should equal the average drop of previous months (or equivalently that the downwards trend in the sales figures should be linear). If the latter decision were made, the problem could have been classified as *Formulate*.

Given the somewhat involved problem analysis above, it was surprising to find that PM918Q05 Charts Question 5 was an easy item, with about 70 % of students correct at the field trial and the main study. Statistically the item behaved extremely well. The students with the correct answer B (370) had the highest ability on all other items, the approximately 20 % of students with answer C (670) had a lower ability overall, and the approximately 5 % answering each of A (70) and D (1,340) had much lower ability again. These good item statistics indicate that the multiple-choice format is working well: students are using their mathematical literacy proficiency to choose the alternative. But what part of this proficiency is most critical? The most common wrong answer was C (670), which is very close to the sales in June. Students giving this answer probably do not have a mathematical concept of 'trend'. Probably they have interpreted "same negative trend" as just a continuation of the same bad sales situation, and not even looked for the decreasing data series. This is a failure related to *Formulate*, not to *Employ*. Amongst students who had a more mathematical concept of trend, the high success rate indicates that many of them were probably able to select answer B (370) on qualitative rather than quantitative grounds. Two choices, B (370) and D (70) were below the June sales figures; choosing B over D is likely to have been supported by reasoning along the lines described above, but done much less precisely without much cognitive demand on the *Employ* process. In summary, it is likely that the major cognitive demand in this item has arisen in *Formulate* and not in the allocated *Employ*. This is a speculative argument based on an interpretation of the item statistics, but it indicates some of the difficulties that can arise in allocating items to just one of the three mathematical processes. In-depth exploration of item performance from this point of view, using the publicly available PISA 2012 international data base, may prove fruitful in understanding items better, and for research.

Using Reported Measures of Mathematical Processes in Teaching

Reporting PISA results by these processes of mathematical literacy may assist educational jurisdictions to review curriculum and teaching. For example a country that has low scores on the *Formulate* process might decide to emphasise this process more in schools, especially by more often beginning with problems in context that need some substantial formulation. This will also involve class discussion about how an element of the real-world context is best described in mathematical terms (e.g. value for money in M154 Pizzas). Teachers may explicitly consider teaching strategies that help students identify mathematical structure and connect problem elements such as the Singapore model method (Fong 1994). A focus on formulation will also involve problems where the solver has to identify multiple relationships (complex or simple) and decide how to put them together, as in PM942Q02 Climbing Mount Fuji Question 2 discussed in Chaps. 4 and 8 of this volume. Teachers can discuss the assumptions behind the models that are used. Even the simplest word problems involve assumptions that are usefully discussed with students and doing this alerts students to how this is essential for applying mathematics. This process can be used to make seemingly unauthentic word problems more realistic. With the pizza problem, students could discuss the assumption that pizzas are circular, the assumption that it is the area of pizza to eat that matters, and how the solution would be modified to find value for money of liquorice strips given their length or value for money of oranges used for juice given their diameters. Research into the teaching of mathematical modelling (see, for example, Blum 2011; Blum et al. 2007) gives many more suggestions. In Chap. 11 of this volume Ikeda shows how using PISA items that focus on particular aspects of the modelling cycle (such as the formulating aspect) can be useful for teaching. Zulkardi in Chap. 15 of this volume describes the creation of PISA-like tasks which reflect life in Indonesia. There is now a big bank of released PISA items to inspire such efforts (e.g. OECD 2013a).

There is no claim that PISA is a full assessment of mathematical modelling. As is evident from the large body of educational research on modelling and applications (e.g. Blum et al. 2007) both teaching and assessment require students to engage with extended tasks even involving multiple trips around the modelling cycle. Along with many other authors, this point is made by Turner (2007) in his presentation of PISA problems with rich classroom potential. Extended tasks can share the PISA philosophy, but they can move considerably away from the PISA format. This is because PISA items must be exceptionally robust. As discussed in Chaps. 6 (by Turner), 7 (by Tout and Spithill) and 9 (by Sułowska) of this volume, they must be suitable for translation into many languages, appropriate for students in many cultures, involve mathematical concepts and processes that are likely to be familiar to students around the world, be able to be consistently scored by many separate teams of markers in an efficient manner, be able to be completed by students within a tight timeframe, have psychometric properties that fit the measurement model well, be self-contained and require very few resources for completion. However, outside of these constraints, many more possibilities exist for designing tasks for teaching and assessing mathematical literacy in a richer way.

In his review of large scale assessment, de Lange (2007) cites initiatives from around the world that assess modelling more completely. Frejd (2013) in an extensive review of the impressive array of recent work compares frameworks and atomistic with holistic approaches. The article recommends that an elaborated judgement of the mathematical and realistic quality of the models produced is required for classroom assessment to improve.

Modelling and Mathematisation Within Mathematics Education

This section aims to clarify the two terms ‘mathematisation’ and ‘modelling’, which readers of the PISA Mathematics Framework will observe have been used with both the same and different meanings at various stages (see also Chaps. 2 and 4 of this volume). They also have various meanings within the broader field of mathematics education. This section exposes and explains these different meanings.

Modelling

Within mathematics education, Kaiser and Sriraman (2006) point out how the term 'modelling' is applied in multiple ways with various epistemological backgrounds to curriculum, teaching and classroom activities. At one end of the spectrum is realistic or applied modelling, and PISA belongs here. This endeavour is dominated by pragmatic goals of solving real-world problems and gaining understanding of the real world. Applied modelling in education was given early prominence by Henry Pollak's survey lecture at ICME-3 in 1976 (Pollak 1979; Blum et al. 2007). Also related to PISA's philosophy through its literacy focus is modelling used for socio-critical goals, with an emancipatory perspective achieved through the capacity to better deal with and understand the world (see also Chap. 1 in this volume). Blum and Niss (1991) point out some of the varying goals and emphases within this tradition of applications and modelling.

At the other end of the spectrum lies what Kaiser and Sriraman (2006) call educational modelling. Here modelling serves the educational goals of developing mathematical theory and fostering the understanding of concepts by starting with real-world situations. The Realistic Mathematics Education tradition at the Freudenthal Institute is the prime example of this approach. Real-world situations are carefully selected to become the central focus for the structuring of teaching and learning a topic, and they provide for students what are now often called 'models of' and 'models for' mathematical concepts that students can use in a process of guided re-invention of mathematics (Gravemeijer and Stephan 2002). The real-world phenomenon models the abstract construction, rather than vice versa as in applied modelling. Classroom materials from the Freudenthal tradition provide many examples of this 'conceptual mathematisation'. For example, de Lange (1987) explains how a situation of aquatic plants growing over a pond, simplified so that the area is doubling every day, can be used to introduce logarithms to students. He defines the base 2 logarithm of a number n to be the time taken for 1 square metre of plants to grow to n square metres. From this definition, students can be guided to discover that the logarithm of 16 is 4 (because the area goes successively from 1 to 2 to 4 to 8 to 16 over 4 days) and can generalise this property. They can also discover the addition law for logarithms. For example, they can discover that $log\ 5 + log\ 7 = log\ 35$ because the plants grow from 1 square metre to 5 square metres in $log\ 5$ days and in the next $log\ 7$ days they grow by another factor of 7. The other properties of logarithms can be deduced in this way, using the real situation as a model for the mathematical theory.

In summary, within the mathematics educational world, modelling is used in multiple senses, which reflect different goals and purposes for using real-world situations in teaching. At one end of the spectrum, which lies entirely within schools, knowledge of the real-world situation is exploited to teach mathematics. The real world 'models' the mathematical world. At the PISA end of the spectrum, lying inside and outside schools, knowledge of abstract mathematics is exploited to better understand the real world. The mathematical world models the real world.

Within schools, the modelling goes in both directions. For nearly everyone, in life beyond school, there is only one direction and that is reflected in the approach taken in PISA. One of the arguments for educational modelling is that it better equips students for applied modelling by contributing “significantly to both the meaningfulness and usability of mathematical ideas” (de Lange 1987, p. 43) and consequently many educational projects include both educational and applied modelling (e.g. Garfunkel’s work in the Consortium for Mathematics and Its Applications COMAP).

Mathematisation in PISA and Elsewhere

The term ‘mathematisation’ has regularly been used in PISA Frameworks. In the Frameworks of 2003, 2006 and 2009 (OECD 2004, 2006a, 2010) it is used to mean the key process behind the Framework (which is called mathematical modelling in the 2012 Framework, aligning it more closely with international usage). In the 2012 Framework *mathematisation* labels the fundamental mathematical capability of moving in either direction between the real world and the mathematical world. In previous PISA Frameworks this was labelled the modelling competency, sometimes with a broader meaning. The translation back to real-world terms is also sometimes called de-mathematising (e.g. OECD 1999, p. 43). These changes have arisen because PISA is a collaboration involving people from different scholarly and educational traditions who use different natural and technical languages to describe what they do. These terminology changes are also discussed by Niss in Chap. 2 and in Chap. 4 by Turner, Blum and Niss in this volume. The present chapter uses the PISA 2012 terminology.

Within the Freudenthal Realistic Mathematics Education tradition, the term ‘mathematisation’ has a central role, referring to a very broad process by which the real world comes to be viewed through mathematical lenses. Mathematics is created in this human endeavour, with the overarching purpose of explaining the world and thereby giving humanity some measure of control over it. This is a philosophical position on the nature and origin of mathematics, as well as a principle guiding teaching. Mathematisation can happen ‘locally’, when a mathematical model for solving a specific problem is created or ‘globally’ for developing a mathematical theory (e.g. logarithms as above) or to tie theories together. It also refers to the process of guided re-invention, when a carefully selected real-world context is used in teaching.

Researchers working within this tradition also distinguish horizontal mathematising which works between reality and mathematics, in both directions, and vertical mathematising where working within the mathematical world provides solutions to problems (locally) or globally develops theory (e.g. generalising logarithms and deducing theorems about them). In Figs. 3.2 and 3.3, horizontal mathematisation in a local situation is depicted by the two horizontal arrows, and vertical mathematisation is depicted by the one vertical arrow in the mathematical world. However, RME's 'global' meaning of mathematisation goes considerably beyond its use in any PISA Framework. In mathematisation, a real-world context can be the inspiration for a mathematical theory or an application of it, or both.

Setting PISA Items in Real World Contexts

Real-world contexts have been at the heart of the mathematical modelling and mathematical literacy discussed above. This section draws on the PISA experience and also the research literature as the specific focus moves from mathematical modelling and turns to some of the challenges that arise from the decision to set (almost) all PISA items in real-world contexts.

The word 'context' is used in several ways in describing educational assessment. Frequently 'context' refers to the conditions under which the student operates. These range from very broad features (e.g. the type of school and facilities), through specific aspects applying to all students (e.g. the purpose of the assessment, done by groups or individuals, timed or not) to the very individual (e.g. this student had a headache). Within PISA mathematics, however, 'context' (and alternatively 'situation') refers specifically to those aspects of the real world that are used in the item. In mathematics education, this is sometimes called the 'figurative context', or the 'objective figurative context' contrasting with the 'subjective figurative context' which refers to the individual's own personal interpretation of that real-world situation. For M154 Pizzas the context includes all the aspects of purchasing pizzas (e.g. that they are a round food, with the most delicious part only on the top), and also more general aspects of shopping including the concept of value for money (which is mathematised as a rate).

Roles of Context in the Solution Process

Knowledge of context can impinge on solutions in many ways. PISA's approach follows that of de Lange (1987). There is a graduation in the importance of the context in solving PISA items. At the lowest level is a unit such as PM918 Charts (see Fig. 3.4) which, as noted above, could have been set in many different contexts with minimal change. This is not to say the context is fully irrelevant to the students' endeavours, even at this lowest level. For a student to feel that they understand the question requires generic real-world knowledge such as why bands are associated with CDs, recognising the abbreviated months of the year, and appreciating that no one is actually kicking kangaroos. Even though knowledge like this does not seem to contribute, students do not do well when they do not understand the basic premises of an item. I recall a boy who told me he could not do a word problem because he did not know what a 'Georgina' was—this girl's name written in the problem was irrelevant to the solution but it stopped him making

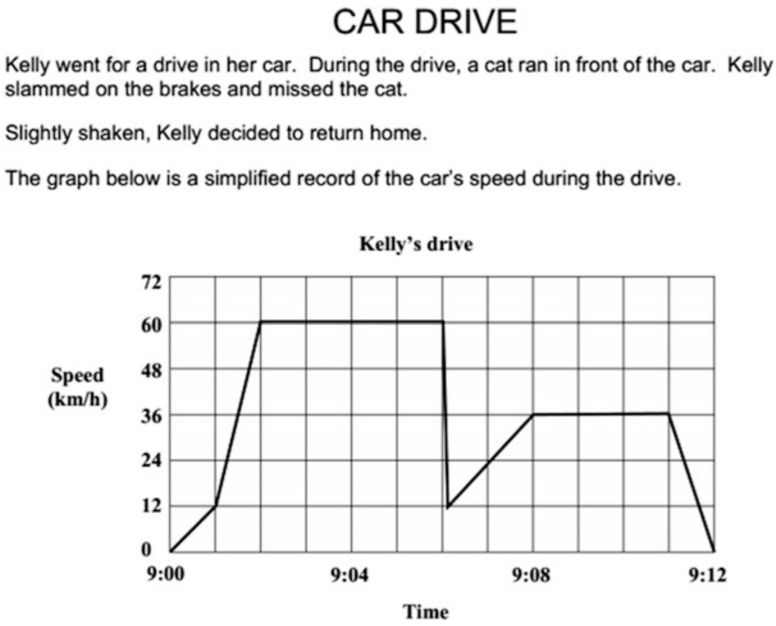


Fig. 3.6 Stimulus for PISA 2006 unit M302 Car drive (OECD 2006b)

progress. As discussed elsewhere, an attractive context may also encourage students to try harder to solve the problem.

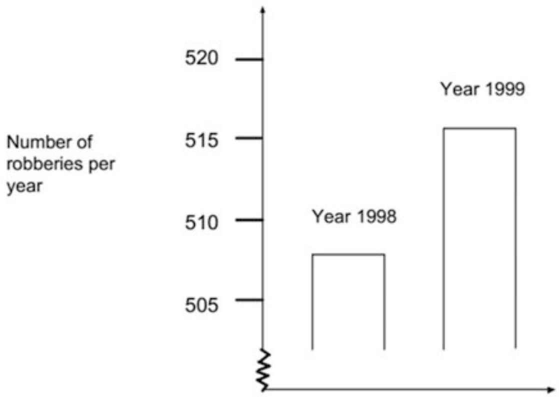
The next level of context use is common in PISA items, where specific features of the context need to be considered in deriving the solution. Appropriate rounding of numbers is frequent e.g. to answer with a whole number of discrete objects (e.g. see PM977Q02 DVD rental Question 2 in Chap. 9 of this volume). The PISA 2006 item M302Q02 Car drive Question 2 (see Fig. 3.6) asked students to give the time when Kelly braked to miss the cat. This requires making the real-world link between braking and decreasing speed, and identifying this feature on the graph.

In a few PISA items, students have to bring into their solutions quite specific real-world knowledge. For example, in the item M552 Rock concert from the field trial for PISA 2003 (OECD 2006a, 2009a) students were given the dimensions of a space for a rock concert, and asked to estimate the number of people it could hold when full with all fans standing. This item required students to make their own estimate of the amount of space that a person would take up in such a concert—information that was not supplied in the item. This has been described as 'second order' use of context (de Lange 1987). Another example of this higher demand of involvement with the context, this time involving *Interpret*, is from the item M179 Robberies (OECD 2006a, 2009a) where students have to comment on an interpretation of a truncated column graph, as shown in Fig. 3.7. Both avoiding the visual trap arising from the truncated columns, and deciding whether the increase should be regarded as large or not, depend on mathematical ability. In essence, this is the

Question 1: ROBBERIES *M179Q01- 01 02 03 04 11 12 21 22 23 99*

A TV reporter showed this graph and said:

"The graph shows that there is a huge increase in the number of robberies from 1998 to 1999."



Do you consider the reporter's statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.

Fig. 3.7 M179 Robberies (OECD 2006b)

ability to see the relevance of both the absolute change and the relative change. Beyond this, the answer also depends on real-world judgements about the robberies context (Almuna Salgado 2010). There would be very different considerations if the graph referred to the number of students attending a school, or the number of parts per million of a toxic chemical in drinking water. Lindenskov in Chap. 15 reports some Danish students' responses to this item.

Measuring students' capacity to solve problems with second order use of context is valuable because it is rare that all the data required is given clearly in a problem in real life. In solving M552 Rock concert, PISA students needed to make an estimate based on body size and personal experience. Outside of the test situation, a real life concert organiser needs to recognise the risks of high crowd density and find published guidelines on crowd safety. In both cases, the problem solver must identify what further information is needed and then access the best available source.

Achieving Authenticity of Context

The definition of mathematical literacy requires that the items used in PISA are authentic: as far as possible they should present students with the challenge of using mathematics in the way in which it is likely to be used in life outside school. Moreover, items should not just *be* authentic; they should *appear to be* authentic so that students feel they are engaged in a sensible endeavour. PISA item writers and selectors give this a high priority so this is one of the criteria on which all countries rate the suitability of items. As is evident from the reports in Chaps. 13, 14 and 15 of this volume, this focus on authentic items has been an important contribution of PISA to mathematics teaching in some countries, which have used items as a model for redesigning school tasks.

Achieving authenticity in items is a complex endeavour. Palm (2006) has created a framework for the characteristics that make a school task authentic. The *event* should be likely to happen and the *question* posed should concord with the corresponding out-of-school situation. The *purpos*e of finding a solution needs to be as clear as it would be in the real situation. The language use (e.g. terminology, sentence structure etc.) should match that used in the real situation. The *information and data* given in the question should be of the type available in the real situation, and the numbers should be realistic. Students should be able to use methods that are available in the real-life setting, not just particular school content, the validity of solutions should be judged against real-world criteria, and the circumstances of performing the tasks (e.g. with calculators) should mimic the real situation. Because PISA attends to these features, it is likely that the item style maximises the chances that students will respond in a realistic way. Many genuine situations are used, such as those in the unit PM923Q03 Sailing ships (see Chap. 1 of this volume). M154 Pizzas gains authenticity by giving the diameter of the pizzas, which I often see alongside prices on the menus in pizzeria. Of course, authenticity is curtailed in an international assessment. One small example is that prices in M154 Pizzas are in the fictional currency of PISA's fictional country Zedland because using realistic prices in the many different currencies around the world would introduce a myriad of variations in the computational difficulty of items. Chapter 7 in this volume gives further examples of this issue.

Palm (2008) provides some evidence that students are indeed more likely to attend to the real-world aspects of the situation when word problems give more details of the situation and attend to the aspects above, although a well-designed study by De Bock et al. (2003) showed that increasing the authenticity of the context by using videos in fact reduced students' success in choosing of sensible models that reflected the real-world situation faithfully. They concluded students may not have expected to process the video information deeply. This is one of many instances where further research would be informative.

Palm's framework has been developed to guide attempts to make school tasks more authentic, and to investigate the well-known phenomenon of students not using their real-world knowledge sensibly within school mathematics. There are many studies that document this, from countries around the world, using word problems such as the one following, where less than 20 % of student solutions were judged realistic:

> Steve has bought 4 planks of 2.5 m each. How many planks of 1 m can he get out of these planks? (Verschaffel et al. 1994, p. 276)

Verschaffel et al. (2009) examine this phenomenon from many points of view. They show how unrealistic problems are a long standing feature of school, by giving historical examples of unrealistic word problems parodied by Lewis Carroll and Gustave Flaubert. From a socio-cultural point of view, students' lack of sense making is in part a reaction to this divorce of school from real life. However, it is also a result of students' superficial mathematisation of the real situations presented even in simple word problems. The extensive series of studies reported in Verschaffel et al. (2009) provide guidance on improving the authenticity of school mathematics even when using simple word problems. Greater effects are likely to come from incorporating realistic modelling into school mathematics, but this is a larger challenge. Studies such as that by Stillman and Galbraith (1998) analyse the ways in which students can be assisted to deal with the cognitive and metacognitive aspects of such complex problems.

It is easy to criticise test items as not being authentic. A salutary experience happened to me many years ago. Some children came home from school and saw the quarterly telephone bill lying on the table. They were shocked to see that the bill was for what seemed to them to be an enormous amount of money. Simplifying the situation, I explained that we had to pay some money to have the telephone and then a certain amount for each call. I intended to leave the discussion there, but the 10 year old wondered aloud how many phone calls the family must have made each day and the children then speculated amongst themselves about this. Shortly after, I wrote a problem for some experimental lessons with the same data and asked 'how many calls per day'. In his feedback, I was surprised to see that the teacher commented especially on this one problem, lamenting the fact that mathematics was full of unrealistic problems that did not interest students, and commented that no child would ever want to know this. Just as a flower withers after it has been picked, a real-world problem often does not stay alive when it is written down on paper. If the techniques adopted by PISA item writers (see Chap. 7 in this volume) are more successful in creating 'face authenticity' of items for students, they could be used in classroom instruction to good effect.

Using Contexts for Motivation

In mathematics teaching, contexts are used for multiple reasons. They are essential to teach students to apply what they learn, and as discussed earlier in this chapter the conceptual mathematisation of a real problem can be used for students to re-invent mathematics through educational modelling. Many teachers also believe that contexts can create positive affect and hence stimulate students' effort to learn and solve problems. Students' genuine interest in a real-world context such as a sustainability issue or the direct relevance of a context to students' lives (e.g. planning a school event) can be harnessed to increase motivation (see, for example, Blum and Niss 1989). Additionally, attractive contexts are very often used simply to enhance the image of mathematics, which some people think is dull, by associating it with pleasurable things (Pierce and Stacey 2006).

Within PISA, contexts are used because doing so is inherent in the definition of mathematical literacy, but there is also a hope that careful choice of contexts that are attractive to 15-year-olds may increase motivation to work at the items. For example, the mathematical core of the unit PM918 Charts could have been tested in many different contexts, so the choice of music bands is likely to have been influenced by the interests of the intended audience of 15-year-olds. Beyond the use of attractive contexts to increase motivation, major issues with the use of contexts are their authenticity (discussed above) and their equity, which is discussed below.

PISA's approach to ensuring the items are as attractive, as equitable and as authentic as possible is three pronged (see also Chaps. 6 and 7 in this volume).

1. Expert opinion on authenticity, interest (and hence motivation) and the equity factors (familiarity and relevance including to subgroups) is sought on each item from every country. Countries also report any cultural concerns to ensure that items do not touch on contexts that are considered inappropriate for use in schools (e.g. gambling, contexts that are potentially distressing).
2. The items use many different contexts and are balanced across the four context categories (*Personal*, *Societal*, *Occupational*, *Scientific*) to minimise the chance of systematic bias arising from the particular contexts chosen.
3. Empirical data from the field trial are used to eliminate from the main survey those items that are easier or harder than expected in some countries, or that show a large gender difference because in these items factors of familiarity or interest or relevance may be differentially affecting performance. One of the reasons for the large item pool taken to the field trial is to allow for this culling. The final findings of overall gender differences are made more robust because the main survey includes only items that did not show large gender differences.

Ensuring Equity

The construction of PISA items must ensure that the survey provides a valid measure of mathematical literacy across countries and groups of students within countries. This is a demanding condition. The use of contexts is essential to PISA, yet it is known that individual students will bring differing background knowledge, interpretations and experiences into the solving process. These differences will affect the survey results when they systematically affect countries or subgroups of interest. Because PISA is not concerned with assessment outcomes of individual students but pools their results, it is not important that every item is fair to every student (that would be impossible) but it is important that, as a whole, every reported group of items is fair to all the targeted groups of students.

Several broad aspects of problems in context are likely to affect an equitable assessment of mathematical literacy: reading demands, the cultural and individual familiarity of the contexts and students' interest in the context. High reading demand was a criticism of early PISA problems, and so attention has been given to simplifying the reading in later surveys. In Chap. 7 of this volume, Tout and Spithill describe some of the rules that are followed. Some strategies for reducing the reading demand reduce authenticity. For example, it is somewhat artificial to provide information question by question as it is required, rather than all together in the stimulus material for a unit. Such competing demands have to be weighed according to their likely effect on the assessment as a whole.

It is clear that the contexts used in PISA must be familiar to the students, at least in the sense that a short text can provide enough information to have students feel confident that they understand the question. In a well-designed study Chipman et al. (1991) found a very small positive effect of context familiarity on word problem performance, with unfamiliarity promoting omission. For tackling PM995 Revolving Door, having seen a revolving door probably gives a small advantage, especially in the initial stages of making sense of the diagrams. However, not everyone who uses a revolving door appreciates how the design blocks the flow of air, and this fact may explain why field trial results did not show differential performance between countries where these doors might be common or not (beyond that predicted by their performance on the item set as a whole).

Critical to PISA is the potential effect of differential familiarity and interest of problem context on performance of countries (addressed through the ratings by each country) and on the subgroups of students for which results are reported such as girls and boys. The research on this is not conclusive. One very frequently cited small scale study is by Boaler (1994), who reported that girls were more likely than boys to be distracted by elements of a context in which they were interested and hence not perform so well. Low and Over (1993) found that girls were more likely than boys to incorporate irrelevant information into solutions (regardless of their interest in the context), although this finding may be an artefact of teaching since the boys and girls were from different (single-sex) schools. On the other hand, the large study by Chipman et al. (1991) found no effect on performance of using problems stereotyped as interesting and familiar to the same or opposite gender or designed to be gender neutral. Familiarity (separately measured) assisted both genders. A recent Dutch study (Hickendorff 2013) of over 600 children found no differential effect of using problems in context for either gender or language ability. This study also found no difference in difficulty between 'naked number' items and word problems, which the author attributed to the Realistic Mathematics Education curriculum in the Netherlands having developed in students a good ability to model real situations. For the purposes of PISA's assessment of mathematical literacy, it is not important whether students perform better or worse on problems in context than on 'naked number' problems, which is what has concerned some researchers. Instead what is important for PISA is that choice of context does not systematically affect the performance of identified groups of students. There are some studies such as that by Cooper and Dunne (1998) that show social class can influence how students work with problems in context, with students of lower social class more likely to draw on their real-world knowledge than the mathematical information specified in the problem statement. If this is a general effect that reflects a difference in ability to use mathematics in context, then it is important that PISA measures it. If it is an artefact of the artificial setting of the assessment, research is needed to eliminate it. We do not know.

Knowledge of the findings of individual studies (rather than the body of evidence) and an acute awareness of the great variety of interests and life experiences around the world have stimulated some critiques of the use of context in PISA problems and claims that a meaningful international assessment using problems in context is impossible. de Lange (2007) reviews these and concludes

> Authors also get quite excited about the role of contexts in large-scale assessments. There are many good reasons to do so, as we still fail to understand quite often the actual role the context plays in a certain problem. . . .. And I would like to add: we cannot say anything firm about the relationship 'context familiarity' to 'success rate'. (p. 1119)

If there are real differences in the mathematical literacy of the targeted groups, then it is important that PISA identifies them. If the differences are due to particular choices in item construction and do not reflect the mathematical literacy construct, it is important that they are eliminated.

In summary, using real-world contexts in items is essential for PISA but raises some important issues. There is potential to motivate students to work hard solving the problems through using attractive contexts, but there is also potential for introducing biases into the assessment. Expert opinion and statistical testing are used by PISA to minimise this threat. Overall, item writers pay serious attention to the authenticity of PISA items, to give as good a measure as possible of students' proficiency to use mathematics beyond school.

Conclusion

The purpose of this chapter has been to examine the links between mathematics and the real world, as they are evident in PISA's concept of mathematical literacy, and to present relevant research and conceptual frameworks. The use of real-world contexts in the teaching and assessment of mathematics has a long history, especially through the use of word problems, which are frequently lampooned for lacking authenticity and relevance. The movement towards mathematical modelling takes the real context seriously. Within mathematics teaching, mathematical modelling goes well beyond the learning of applied mathematics, where techniques for standard problems in areas of application (such as physics) are taught and practised, aiming to teach students to develop their own mathematical models, and to interpret results in real-world terms, as well as to solve the intra-mathematical problems involved. Mathematical literacy lies within this concept of mathematical modelling. The chapter also discussed the way in which the PISA teams have worked within the strong constraints of an international assessment to develop survey items that use real-world contexts in a way that motivates students to solve the items, and to make these items as equitable as possible taking into account their varying familiarity, interest and relevance to groups of students. The high authenticity of PISA items, especially considering the constraints of the international assessment situation, have provided a model and resources for authentic problem solving in schools that is relatively easy to implement, as well as resources to inspire more extended problem solving.

References

Almuna Salgado, F. J. (2010). *Investigating the effect of item-context on students' performance on mathematics items*. Masters Coursework thesis, Melbourne Graduate School of Education, The University of Melbourne. http://repository.unimelb.edu.au/10187/10041

Blum, W. (2011). Can modelling be taught and learnt? Some answers from empirical research. In G. Kaiser et al. (Eds.), *Trends in teaching and learning of mathematical modelling (ICTMA 14)* (pp. 15–30). Dordrecht: Springer.

Blum, W., & Niss, M. (1989). Mathematical problem solving, modeling, applications, and links to other subjects — State, trends and issues in mathematics instruction. In W. Blum, M. Niss, & I. Huntley (Eds.), *Modelling, applications and applied problem solving* (pp. 1–21). Chichester: Ellis Horwood.

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects. *Educational Studies in Mathematics, 22*, 37–68.

Blum, W., Galbraith, P., Henn, H.-W., & Niss, M. (Eds.). (2007). *Modelling and applications in mathematics education*. New York: Springer.

Boaler, J. (1994). When girls prefer football to fashion? An analysis of female under achievement in relation to realistic mathematics contexts. *British Educational Research Journal, 20*(5), 551–564.

Burkhardt, H. (1981). *The real world and mathematics*. Glasgow: Blackie.

Chipman, S., Marshall, S., & Scott, P. (1991). Content effects on word problem performance: A possible source of test bias? *American Educational Research Journal, 28*(4), 897–915.

Cooper, B., & Dunne, M. (1998). Anyone for tennis? Social class differences in children's responses to national curriculum mathematics testing. *The Sociological Review, 46*(1), 115–148.

Cundy, H. M., & Rollett, A. P. (1954). *Mathematical models*. Oxford: Clarendon Press.

de Bock, D., Verschaffel, L., Janssens, D., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems. *Learning and Instruction, 13*, 441–463.

de Lange, J. (1987). *Mathematics—Insight and meaning*. Utrecht: Rijksuniversiteit Utrecht.

de Lange, J. (2007). Large-scale assessment and mathematics education. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1112–1142). Charlotte: Information Age Publishing.

Fong, H. K. (1994). Bridging the gap between secondary and primary mathematics. *Teaching and Learning, 14*(2), 73–84. http://repository.nie.edu.sg/jspui/bitstream/10497/479/1/TL-14-2-73.pdf. Accessed 3 Dec 2013.

Frejd, P. (2013). Modes of modelling assessment—A literature review. *Educational Studies in Mathematics, 84*, 431–438.
Gravemeijer, K., & Stephan, M. (2002). Emergent models as an instructional heuristic. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 145–169). Dordrecht: Kluwer.
Hickendorff, M. (2013). The effects of presenting multidigit mathematics problems in a realistic context on sixth graders' problem solving. *Cognition and Instruction, 31*(3), 314–344. doi:10.1080/07370008.2013.799167.
Kaiser, G., & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM, 38*(3), 302–310.
Low, R., & Over, R. (1993). Gender differences in solution of algebraic word problems containing irrelevant information. *Journal of Educational Psychology, 85*(2), 331–339.
Organisation for Economic Co-operation and Development (OECD). (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2004). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing. doi:10.1787/9789264101739-en.
Organisation for Economic Co-operation and Development (OECD). (2006a). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD Publishing. doi:10.1787/9789264026407-en.
Organisation for Economic Co-operation and Development (OECD). (2006b). *PISA released items*. http://www.oecd.org/pisa/38709418.pdf. Accessed 3 Dec 2013.
Organisation for Economic Co-operation and Development (OECD). (2009a). *Learning mathematics for life: A perspective from PISA*. Paris: OECD.
Organisation for Economic Co-operation and Development (OECD). (2009b). *PISA: Take the test*. Paris: OECD Publications. http://www.oecd.org/pisa/pisaproducts/Take%20the%20test%20e%20book.pdf. Accessed 17 May 2014.
Organisation for Economic Co-operation and Development (OECD). (2010). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. PISA: OECD Publishing. doi:10.1787/9789264062658-en.
Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en
Organisation for Economic Co-operation and Development (OECD). (2013b). *PISA 2012 released mathematics items*. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf
Ormell, C. P. (1972). Mathematics, applicable versus pure and applied. *International Journal of Mathematical Education in Science and Technology, 3*(2), 125–131. doi:10.1080/0020739700030204.
Palm, T. (2006). Word problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics, 26*(1), 42–47.
Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*(1), 37–58.
Pierce, R., & Stacey, K. (2006). Enhancing the image of mathematics by association with simple pleasures from real world contexts. *Zentralblatt fur Didaktik der Mathematik, 38*(3), 214–225.
Pollak, H. (1979). The interaction between mathematics and other school objects. In UNESCO (Ed.), *New trends in mathematics teaching IV* (pp. 232–248). Paris.
Stillman, G., & Galbraith, P. (1998). Applying mathematics with real world connections: Metacognitive characteristics of secondary students. *Educational Studies in Mathematics, 36*(2), 157–195.
Stillman, G., Galbraith, P., Brown, J., & Edwards, I. (2007). A framework for success in implementing mathematical modelling in the secondary classroom. In J. Watson & K. Beswick (Eds.), *Proceedings of the 30th annual conference of the Mathematics Education*

Research Group of Australasia. MERGA, http://www.merga.net.au/documents/RP642007.pdf. Accessed 2 Dec 2013.

Synge, J. L. (1951). *Science: Sense and nonsense*. London: John Cape.

Turner, R. (2007). Modelling and applications in PISA. In W. Blum, P. L. Galbraith, H. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (The 14th ICMI study, pp. 433–440). New York: Springer.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 273–294.

Verschaffel, L., Greer, B., van Dooren, W., & Mukhopadhyay, S. (Eds.). (2009). *Words and worlds: Modelling verbal descriptions of situations*. Rotterdam: Sense Publishers.

Wikipedia. Mathematical model. http://en.wikipedia.org/wiki/Mathematical_model. Accessed 18 Nov 2013.

Chapter 4
Using Competencies to Explain Mathematical Item Demand: A Work in Progress

Ross Turner, Werner Blum, and Mogens Niss

Abstract This chapter describes theoretical and practical issues associated with the development and use of a rating scheme for the purpose of analysing mathematical problems—specifically, to assess the extent to which solving those problems calls for the activation of a particular set of mathematical competencies. The competencies targeted through the scheme are based on the mathematical competencies that have underpinned each of the PISA Mathematics Frameworks. The scheme consists of operational definitions of the six competencies (labelled as *communication; devising strategies; mathematisation; representation; using symbols, operations and formal language;* and *reasoning and argument*), descriptions of four levels of activation of each competency, and examples of the ratings given to particular items together with commentary that explains how each proposed rating is justified in relation to the competency definition and level descriptions. The mathematical problems used so far to investigate the action of those competencies are questions developed for use in the PISA survey instruments from 2000 through to 2012. Ratings according to the scheme predict a large proportion of the variation in difficulty across items, providing evidence that these competencies are important elements of students' problem solving capabilities. The appendix gives definitions of each competence and the specification of each of four levels for each.

R. Turner (✉)
International Surveys, Educational Monitoring and Research, Australian Council for Educational Research (ACER), 19 Prospect Hill Rd, Camberwell, VIC 3124, Australia
e-mail: Ross.Turner@acer.edu.au

W. Blum
Institute of Mathematics, University of Kassel, FB10 Mathematik and Naturwissenschaften, Heinrich-Plett-Str 40, 34132, Kassel, Germany
e-mail: blum@mathematik.uni-kassel.de

M. Niss
IMFUFA/NSM, Roskilde University, Universitetsvej 1, Bldg. 27, 4000, Roskilde, Denmark
e-mail: mn@ruc.dk

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_4

Introduction

In Chap. 2 of this volume, Mogens Niss describes a set of competencies that have been central to the definition of mathematical literacy within the PISA context, and have been increasingly instrumental to the design of PISA mathematics items. Indeed Niss's work outlines what might be referred to as a 'competence model' of mathematical proficiency, in which proficiency can be seen as a function of the extent to which an individual possesses and is able to mobilise certain mathematical competencies. Investigative work described in the present chapter shows how these competencies can help to understand the cognitive demand and predict the empirical difficulty of PISA mathematics items. This in turn suggests that the competencies form a very important part of the cognitive actions taking place when individuals attempt to solve certain types of mathematical problems. That kind of knowledge has also been of assistance to test item developers, by helping them in targeting their development work more efficiently. It is also likely to be of relevance to mathematics teachers as they design teaching and learning activities to improve the mathematical proficiency of their students. This chapter describes the development and key features of a scheme for evaluating PISA test items according to the extent to which the processes of solving the problems demand activation of the mathematical competencies (called the fundamental mathematical capabilities in the PISA 2012 Framework). The development work is ongoing; nevertheless use of the scheme has already borne fruit.

Background and Context

The processes and outcomes of survey instrument development, survey implementation, data generation, and data analysis associated with the PISA survey have presented many opportunities for participating countries and others involved in PISA to investigate a wide variety of educational and technical matters. The PISA Mathematics Expert Group (subsequently referred to as the MEG) in October 2003 began an investigation of the PISA items that had been developed for use in the 2003 survey when mathematics first took its place as the major PISA test domain. Initially, the focus of that investigation was on aspects of item and test validity that had been raised a year previously as an issue requiring attention in the item development process. Several questions were posed by the MEG members as part of its process of test item development. To what extent did the test items under development reflect the Framework? To what extent did the items give an indication of mathematical literacy? Would the PISA measure of mathematical literacy be confirmed through other tests of mathematical literacy? Do PISA results predict something about later levels of mathematical proficiency, for example adult mathematical literacy?

The posing of those questions led to some concentrated work by different members of the MEG to investigate aspects of the validity of the PISA mathematics test instrument then under development. One direction in particular lay in examining factors related to the empirical difficulty of PISA mathematics items. In an unpublished discussion paper developed on this topic, Blum and de Lange noted that while certain factors that make mathematics items more or less difficult could not be easily investigated in a large-scale study such as PISA (in particular, personal factors such as "individual pre-knowledge or individual motivations/emotions"), what could be investigated is

> . . . on the one hand, to describe as precisely as possible certain external features of items as well as the cognitive demands that items impose on the problem solver and, on the other hand, to establish statistical correlations between characteristics of items and the empirical item difficulty (in the whole population). This can be done by methods such as regression analysis.

They also noted that

> In order to describe cognitive demands of items one needs to have at one's disposal appropriate "competence models" (like the one we have developed for PISA mathematics). Then one has to compile, for each item, ideal typical solution processes and to identify those "competence elements" (knowledge & skills, images/"Vorstellungen", abilities/competencies) that have to be activated during these processes, including the cognitive level of this activation. If one distinguishes for each competency (for instance: mathematical argumentation) let's say three levels (0— not necessary, 1—moderately necessary, 2—substantially necessary) then for each item and each competency there is a certain number (describing the cognitive level of activation of this competency for solving this item). (Blum and de Lange, unpublished MEG meeting document, October 2003)

This discussion set the scene for an investigation of the relationship between the competence model underpinning the PISA Mathematics Framework on one hand (some set of competencies underpin mathematical literacy, and those competencies need to be activated by individuals in order for them to solve mathematical problems), and the empirical difficulty of PISA mathematics test items on the other. The central question posed in designing the investigation was whether and how the mathematical competencies needed to solve PISA problems were connected to the empirical difficulty of the problems. Two kinds of connection were envisaged. First, if solving one problem requires drawing on a wider range of competencies than solving another problem, how would that difference be reflected in the relative difficulty of the two problems? Second, to the extent that different levels of activation of a particular competency could be identified (for example no activation at all, activation to a small degree, activation to a large degree) would the *degree* of activation of competencies required for a particular problem be related to the difficulty of that problem?

Blum and de Lange argued that the choice of variables to use in such an investigation should largely be a theoretical matter, and proposed as a starting point considering a set of variables that flowed out of task analysis work done previously in the German context in the COACTIV project (Neubrand et al. 2013),

Table 4.1 Initial set of variables proposed for item difficulty research (Blum and de Lange 2003)

	Variable	Possible level definitions
Surface features of items	1. Mathematical topic	1 Arithmetic, 2 algebra, 3 geometry, 4 probability and statistics
	2. Overarching idea	1 Quantity, 2 change and relationships, 3 space and shape, 4 uncertainty
	3. Item format type	1 Multiple choice, 2 closed constructed, 3 open constructed
	4. Context type	0 zero, 1 intra-mathematical, 2 quantities, 3 close to reality, 4 authentic
Cognitive demand characteristics of items	5. Concept images (“Grundvorstellungen”) needed	0 none, 1 only one elementary, 2 several elementary or one non-elementary, 3 more
	6. Extent of solution process	1 only one step, 2 two or three steps, 3 more
	7. Argumentation competency needed	0 none, 1 moderate, 2 substantial
	8. Modelling competency needed	0 none, 1 moderate, 2 substantial
	9. Communication competency needed	0 none, 1 moderate, 2 substantial
	10–14 . . . See remaining PISA competencies	0 none, 1 moderate, 2 substantial

combining surface features of the items and several cognitive characteristics. The set of variables proposed for consideration at that time are presented in Table 4.1.

The conception of the investigation planned at that time was to identify a set of factors, or variables, which would be a mixture of surface features and cognitive dimensions, and to rate items according to the applicable characteristics and the demand for activation of the cognitive dimensions as part of the solution process, resulting in a several-dimensional vector that would describe important aspects of the cognitive demand of each item. The terminology of *item demand* was established as a reference to the number and nature of aspects of the item that were called in to play as part of the solution process and the level at which the aspects were called in to play. There was a clear expectation that the process of examining and assigning ratings to items would lead to further consideration of the competence model being used, and an iterative process of refinement and development would likely ensue. Indeed that is exactly what has occurred.

Features of an Item Analysis Scheme

Following that initial discussion among members of the MEG, a research team comprising some members of the MEG continued to develop and refine a scheme for evaluating mathematics problems. The main objective was to better understand the drivers of item difficulty. The scheme would consist of a set of variables, operational definitions of those variables, and descriptions of levels within each variable. The particular variables chosen for investigation arose from the PISA competence model referred to earlier.

In the original proposal, several item characteristics had been suggested as variables alongside mathematical competencies. Specifically, as shown in Table 4.1, inclusion of information about surface features of mathematics tasks such as the question format, content category, context type, or mathematical topic area had been proposed. However, the process of developing and selecting items for use in the PISA survey instruments involved consciously balancing several of those surface factors with respect to item difficulty as far as was possible (OECD 2003 p. 50). A design objective was to produce items within each category that had as wide a range of difficulties as possible, in order to avoid the unintended possibility that student performance on different items may be systematically affected by factors unrelated to the measured construct. For example items allocated to the four context categories (see Chap. 1, this volume) defined in the Framework need to span the difficulty spectrum, but these categories are not seen as fundamental to the mathematical literacy construct. Similarly, the item developers consciously aimed to have as full a range of difficulties as possible for items presented in each of the item format types (such as multiple-choice format, and open-ended items). For this reason it was not expected that surface characteristics such as these would contribute useful information in the analysis of the relationship between item cognitive demand and empirical item difficulty.

For the purpose of this investigation, the initially proposed variables were reduced to a set of six variables based on a reconfiguration of the 'Niss competencies' that had been a central element of the PISA Mathematics Framework since PISA's inception (for example, as originally articulated for PISA (OECD 1999) and in the most recent Framework (OECD 2013b)). The origin of this set of competencies, and their use and development over several PISA survey administrations, is discussed in detail by Niss in Chap. 2. Using the six competencies and a procedure for assigning ratings to mathematics test items according to the extent to which solving each item calls on activation of each of the defined variables, has generated sets of ratings that have been used as data to examine the relationship between demand for activation of the competencies in solving PISA mathematics items, and the empirical difficulty of those items as measured through the various PISA survey administrations.

Building Competency Definitions and Level Descriptions

The eight mathematical competencies of the first PISA Framework (OECD 1999) provided a starting point for building a scheme to analyse the competency-related demands imposed by the solution processes needed for a range of mathematical tasks. To build a scheme that would be as compact and manageable as possible within the context of an international survey, those eight competencies were reconfigured as six in the initial PISA version of the scheme: *reasoning and argument* (including mathematical thinking, reasoning, argumentation, and justification); *communication*; *modelling*; *representation*; *problem solving*; and *using symbolic, formal and technical language and operations* (abbreviated as 'symbols and formalism'). Thus the two Niss competencies (see Chap. 2) of *mathematical thinking* and *mathematical reasoning* were combined into one, and the *mathematical aids and tools* competency was dropped as being inappropriate in the context of PISA tasks, which at that time were all paper-based. Operational definitions of each of the chosen competencies were devised, together with a description of four levels of activation of each competency. The initial definitions and descriptions are reproduced in Appendix 1.

However, the initial definitions and level descriptions have undergone significant and progressive change over a period of years in which the scheme has been put to use to analyse PISA mathematics tasks. For example, the competency that was initially labelled *modelling* was first defined as "Mathematising, interpreting, validating." Subsequently, the label was changed to *mathematising* and the definition has become "Translating an extra-mathematical situation into a mathematical model, interpreting outcomes from using a model in relation to the problem situation, or validating the adequacy of the model in relation to the problem situation." The following section of this chapter describes the issues thrown up for the investigators to consider as they applied the scheme, generated sets of item ratings, and analysed those ratings.

Two sets of ratings and their statistical analysis have been reported publicly, with both of them providing similar pictures of the strengths and weaknesses of the scheme as it developed during the period in which those two phases of the research were conducted.

The first results were presented at the PISA Research Conference in Kiel, Germany, in 2009 and subsequently published in the Proceedings (Turner et al. 2013). That analysis was based on two sets of ratings of the 48 mathematics items that had been used in both the PISA 2003 and PISA 2006 survey instruments: the first set of ratings provided by eight raters working independently; and the second set being ratings of the same 48 items 2 years later by a different (but overlapping) set of raters, again completing the task independently, and using a scheme that had changed in only very minor ways. The items were rated according to their demand for activation of the six aforementioned competencies, in accordance with competency definitions and descriptions of four possible levels of activation of each of the competencies. The ratings by individual raters for each item were averaged to provide the final rating for each competency for the item.

Analysis of the data showed that a regression model that included the ratings for just three of the six competencies (those labelled *reasoning and argument*, *symbols and formalism*, and *problem solving*) could account for more than 70 % of the variability of item difficulty across this set of 48 items (for details, see Turner et al. 2013).

A reasonable level of consistency was achieved among the different raters, but there was enough evidence of idiosyncrasy on the part of individual raters and inconsistency across the raters in relation to particular items to suggest that the shared understanding of the meaning of the competencies and the standards defined by the level descriptions of each competency could be further enhanced. In particular, discussion among the raters demonstrated that in some cases similar ratings had been assigned for very different reasons, while in other cases assigned ratings were widely divergent. The success of the regression model showed that when ratings were averaged across a small group of raters, useful data were derived, but considerable variability was observed across the raters indicating there was room for further refinement of the scheme. The observation that three of the competencies did not appear to contribute usefully to the prediction model provided some direction as to which of the competency definitions could usefully be revised.

A third set of ratings was reported publicly in late 2011. The ratings had been produced in 2011 by five members of the research team using a revised version of the scheme, independently analysing a total of 196 test items that had been newly developed for possible use in the 2012 PISA survey. The scheme as used in that exercise was described by Turner (2012) and data were analysed as reported by Turner and Adams (2012). From that analysis, a prediction model with reasonably good properties that involved three of the competencies, *devising strategies* (which was a re-named and differently defined version of the former *problem solving* competency), *communication,* and *symbols and formalism*, was shown to account for some 74 % of the variation in difficulty of those test items, an even higher proportion than in the first published analysis.

The analysis showed that significant overlap existed between the newly defined variables *devising strategies, mathematising,* and *reasoning and argument,* so only one of these was included in the prediction model. The *communication* competency now seemed to be contributing usefully to the prediction (whereas it had not in the earlier analysis) and the *symbols and formalism* competency continued to contribute. Two of the competencies, *representation* and *mathematisation,* were found in both sets of analysis not to contribute useful information to the prediction model (in the case of *mathematisation,* information that was not already captured by other competencies in light of the very high observed correlations). Nevertheless it was evident from the two sets of analyses that adjusting the wording of the definitions, and the descriptions of levels of operation of each competency, had led to a marked change in the way the scheme had functioned, although the good prediction of difficulty was maintained.

While the scheme had been further developed as these rating and analysis exercises went on, the two phases of rating and analysis pointed to several features of the definitions and descriptions as being potentially problematic. A more focused review of the category definitions and level descriptions was instituted as a result.

The first issue was that there was not sufficient agreement on the boundaries between related categories. In the original set of competency definitions (in Appendix 1) the text clearly anticipates overlap, which is consistent with the assumption of overlap in the KOM competency scheme described by Niss in

Chap. 2 of this volume. For example, both the *reasoning and argumentation* and the *representation* competency definitions include the rather confusing statement 'can be part of problem solving process' without in any way attempting to clarify when an item demands *reasoning and argumentation*, or *representation*, and when it demands *problem solving*; nor did it show the relationship between these aspects of cognitive demand, and any implications this relationship might have for the item rating task, thereby leaving the door wide open for different interpretation by raters. Similarly, the *problem solving* definition includes the phrase 'and implementation of the mathematical solution processes, largely within the mathematical world' and that wording opens the way to significant overlap with the *symbols and formalism* competency, and perhaps others. A further example of lack of clarity in the distinctions between competencies is in the original *symbols and formalism* definition, which includes the phrase 'using particular forms of representation…' without clarifying at all where this competency ends and where the *representation* competency begins. The formulations used in that set of definitions did not sufficiently clarify the boundaries among the competencies.

The problem, though, continued to be apparent in the revised descriptions. For example, there was no clear agreement on where the strategic thinking involved in devising a suitable strategy for solving a problem and monitoring its implementation ended, and where processes of mathematical reasoning to solve the problem commenced. It was also clear that the definition of *mathematising* did not support a sufficiently consistent interpretation, so that in some cases one rater may have used the *mathematising* competency while another may have used the *symbols and formalism* competency to describe essentially the same aspect of mathematical thinking and processing, namely setting up a formula or an algebraic expression as a mathematical model of a given real-world situation. It became clear that further work was needed to better delineate the meaning of each the competencies in order to give them operational definitions that identified and were built on separate aspects of each process.

A second problematic feature, closely related to the first, stemmed from the observation that typically more than one of the competencies as they were defined at that time was required to solve a problem. When the activation of several competencies is necessary to solve a problem, as is typically the case in PISA tasks, identifying which one competency is the most important, or which of the competencies are more important than others, proved very difficult, and different raters frequently made different judgements about this. Operation of a 'halo effect' might lead raters to rate an item at a similar level for each relevant competency, for example for very demanding items to assign high ratings to all competencies just because the item seems relatively difficult, or to assign all low ratings for a very straightforward and easy item. This would lead to high correlation between the ratings for each competency and this was likely a major cause of the outcomes of the statistical analysis of the first sets of ratings that showed the best predictive model required only three of the competencies. Whilst the instructions for the rating exercise had recommended that the rater should identify which single competency is the most central to the item, and treat other related competencies by separating out their unique contribution over and above that which is already covered by the 'main' competency, this proved a difficult judgement to make in practice. Indeed some of those involved in the rating exercise questioned whether such a goal could or should be achieved at all. This raised questions about whether six competencies were required or whether perhaps fewer would be sufficient. For example, if it is always or almost always true that *reasoning* and *devising strategies* occur together, perhaps they should be combined into a single more general competency that encompasses them both. This question reinforced the need to further explore separation of the competency definitions. This experience also highlighted the fact that the procedure to be adopted when making the ratings was an essential part of the scheme.

A third feature that appeared to cause difficulties to users of the scheme was the way in which the level of demand for each competency was described. Two aspects of this issue were identified. The first followed from the observation that several of the adjectives used to describe different levels of activation of each competency were rather generic, relative terms that did not convey sufficient objective meaning to different users of the scheme. For example, words like 'simple' and 'complex' did not support consistent interpretation and categorisation of problem solving events, and words like 'familiar' have a curricular or experiential connotation that is counter to the cognitively oriented definitions of the competency levels. Certainly words such as these would tend to mean something very different for students at different stages of their education. The decision was taken to revise the level descriptions to minimise the incidence of unclear adjectives of this kind, and where that proved difficult, to provide further examples in order to clarify the intended meaning of those words. But a second and more fundamental question requiring an answer was just what aspects of each competency change as the level of activation changes. A further aspect of an overhaul of the level descriptions, therefore, was to have a fresh look at what aspects of demand for each competency would most effectively capture gradations in the degree of demand.

Since the first attempt to operationally define the variables and levels, the current authors and their research collaborators have made ongoing attempts to revise the definitions and descriptions in order to reduce the impact of the three issues identified in the preceding paragraphs. Appendix 2 presents the current set of competency definitions and level descriptions, which reflect the progress made to date in the refinement of the scheme in an attempt to address the problems identified through its early uses. The following sections describe how these issues were confronted as the scheme was developed.

One further factor that causes some of the observed variability in the ratings assigned by different raters to particular items is that for some items, different methods of solution may call for the activation of the competencies in a different combination or at different levels. It is recognised, therefore, that some degree of variability in rating outcomes is inevitable. For PISA ratings, the advice was to rate the solution which was judged by the rater to be most likely given by 15-year-old students.

Competency Definitions

For a scheme such as this to work well, there is clearly a need to devise operational definitions of the six competencies that will maximise the distinctions between competencies, and will therefore help users of the scheme to treat particular aspects of problem demand more reliably and consistently. Ideally, when a specific cognitive demand within a solution is identified, the associated competency will be unambiguous and should support consistent ratings. Making these definition is an especially challenging task when, as in PISA, end users will have different languages and education traditions.

In Table 4.2, the development of the operational definition of the *communication* competency is traced as a first example of how the definitions have changed over time. The set of definitions shows the development from the initial version (Appendix 1) to the current version (Appendix 2), and all of the intervening versions. The definitions have become progressively longer as more and more features have been added in an attempt to delineate the competency. From the beginning, this competency included both a receptive and an expressive component. The expressive component expanded early and remained unchanged after that. But the receptive component has continued to change to clarify which elements of the question statement should be taken into account as part of the competency, and towards its main emphasis being on understanding and interpreting the situation presented. This leads to additional descriptive material (in Appendix 2) supporting the 2013 definition that aims to put the focus of the receptive aspect of this competency on understanding what the task asks the problem solver to achieve, and not on the

Table 4.2 Development of *communication* definition

Communication	
2005a	Decoding and interpreting stimulus, question, task; expressing conclusions
2005b	Decoding and interpreting stimulus, question, task; explaining one's work, expressing conclusions
2006	Decoding and interpreting statements, questions and tasks; including making sense of the information provided; presenting and explaining one's work or reasoning
2007	Decoding and interpreting statements, questions and tasks; including imagining the situation presented so as to make sense of the information provided; presenting and explaining one's work or reasoning
2011a	Decoding and interpreting statements, questions, tasks and objects; imagining and understanding the situation presented and making sense of the information provided; presenting and explaining one's mathematical work or reasoning
2011b	Reading, decoding and interpreting statements, questions, tasks and objects; imagining and understanding the situation presented and making sense of the information provided; presenting and explaining one's mathematical work or reasoning
2013	Reading and interpreting statements, questions, instructions, tasks, images and objects; imagining and understanding the situation presented and making sense of the information provided including mathematical terms referred to; presenting and explaining one's mathematical work or reasoning

Table 4.3 Development of the *devising strategies for solving problems* definition

Devising strategies for solving problems (originally labelled 'Problem solving')	
2005	The planning, or strategic controlling, and implementation of mathematical solution processes, largely within the mathematical world
2006	Selecting or creating a mathematical strategy to solve problems arising from the task or context; successfully implementing the strategy
2007	Selecting or devising, as well as implementing, a mathematical strategy to solve problems arising from the task or context
2013	Selecting or devising a mathematical strategy to solve a problem as well as monitoring and controlling implementation of the strategy

Table 4.4 Development of the *mathematising* definition

Mathematising (originally labelled 'modelling')	
2005	Mathematising, interpreting, validating
2006	Mathematising an extra-mathematical situation, or making use of a given or constructed model by interpreting or validating it in relation to the context
2007	Mathematising an extra-mathematical situation (which includes structuring, idealising, making assumptions, building a model), or making use of a given or constructed model by interpreting or validating it in relation to the context
2013	Translating an extra-mathematical situation into a mathematical model, interpreting outcomes from using a model in relation to the problem situation, or validating the adequacy of the model in relation to the problem situation

interpretation of (for example) the mathematical content of any *representations* present. Understanding the goal of the task is an essential precursor to the detailed mathematical thinking and work needed to achieve that goal, with the expectation that the subsequent thinking and work would form part of other competencies.

In Table 4.3, the development of the *devising strategies* competency is traced. The label of the *problem solving* competency was changed to *solving problems mathematically* and then to *devising strategies for solving problems*. This change reflected a shift in emphasis from a focus on the processes and steps of a problem solution to the processes of planning how to go about solving a problem, planning a solution path, and monitoring the implementation of the strategy. The change was intended to help users focus on the strategic thinking required, and therefore help avoid some of the previous overlap particularly with the reasoning, the modelling, and the symbols and formalism activities that flowed from a focus on implementation of the strategy implied by the original label.

The development of the *mathematising* competency definition is tracked in Table 4.4. The term 'modelling' carries certain baggage with it so that in the minds of many people it would include all aspects of the modelling cycle (including the formulating, mathematical processing, interpreting and validating aspects). The changes to the label and to the operational definition here were intended to narrow the focus to the parts of the modelling cycle (see Chap. 3) that are about the direct interface between the context and its mathematical expression, hence to only the

Table 4.5 Development of the *representation* definition

Representation	
2005	Concrete expression of an abstract idea, object or action; a transformation or mapping from one form to another; can be part of modelling or problem solving
2006	Interpreting, translating between, and making use of given representations; selecting or devising representations to solve problems or to present one's work
2007	Interpreting, translating between, and making use of given representations; selecting or devising representations to solve problems or to present one's work. The representations referred to are depictions of mathematical objects or relationships, which include equations, formulae, graphs, tables, diagrams, pictures, textual descriptions, concrete materials
2011	Interpreting, translating between, and making use of given mathematical representations; selecting or devising representations to capture the situation or to present one's work. The representations referred to are depictions of mathematical objects or relationships, which include symbolic or verbal equations or formulae, graphs, tables, diagrams
2013	Decoding, translating between, and making use of given mathematical representations in pursuit of a solution; selecting or devising representations to capture the situation or to present one's work

steps of transforming some feature of the problem context into a mathematical form (the process *Formulate* of Chap. 1) or interpreting mathematical information in relation to the elements of the context it reflects (the process *Interpret* of Chap. 1). The critical defining feature identified in clarifying this competency lies in the active connection of a real-world context with a mathematical expression of some feature of the context. A benefit of this would be to separate the intra-mathematical processing work, the manipulation of mathematical representations, and perhaps the mathematical reasoning elements from the way the *mathematisation* competency should be used within this scheme.

Changes over time to the definition of the *representation* competency are presented in Table 4.5. This competency is one that appears to have contributed little to understanding the drivers of item difficulty, yet it is seen as an important mathematical competency and so arguably should remain in the scheme. The development of the definition shows a number of features and different attempts to resolve potential overlap and confusion in its use. The original definition referred to both modelling and problem solving without attempting to clarify the particular aspects of those activities that should be considered as part of the *representation* competency. The confusion with the *mathematising* variable is also evident in the original level descriptions for *representation* where the relationship between the representation and the feature being represented are prominent. The key elements around which clarification has been sought are the need to include both devising mathematical representations and using given representations, as well as a delineation of which problem elements should be regarded as mathematical representations for the purposes of this scheme. In the explanatory text written to support interpretation of the current version (presented in Appendix 2), the words decoding, devising, and manipulating are included to guide the user to a clearer understanding of what actions are relevant, in addition to the demand of linking different

Table 4.6 Development of the using symbol, operations and formal language definition

Using symbols, operations and formal language	
2005	Activating and using particular forms of representation governed by special rules (e.g. mathematical conventions)
2006	Understanding, manipulating, and making use of symbolic expressions (including using arithmetic expressions and carrying out computations), governed by mathematical conventions and rules; understanding and utilising constructs based on definitions, rules and formal systems
2011	Understanding and implementing mathematical procedures and language (including symbolic expressions and arithmetic operations), governed by mathematical conventions and rules; understanding and utilising constructs based on definitions, rules and formal systems
2013	Understanding and implementing mathematical procedures and language (including symbolic expressions, arithmetical and algebraic operations), using the mathematical conventions and rules that govern them; activating and using knowledge of definitions, results, rules and formal systems

representations to each other; and the list of included mathematical entities is further clarified. In particular, the potential overlap between interpreting and using mathematical representations on the one hand, and the interpretation involved in the *communication* competency is addressed; as is the potential overlap between the use of symbolic forms of representation as part of this competency or part of the *symbols and formalism* competency.

In Table 4.6, the developmental stages of the *using symbols, operations and formal language* competency definition are presented. This competency was originally labelled as *using symbolic, formal and technical language and operations* following the full name used in the PISA Framework, but has generally been referred to as the *symbols and formalism* competency. It has consistently come out as a strong predictor of item difficulty, and is clearly a key element of a competency-based scheme since mostly at least some formal or technical operations have to be carried out in conjunction with other activities in order to solve a mathematical problem. The potential overlap with the *mathematising* competency was addressed through the text that locates the *Formulate* process (including with symbolic expressions) in *mathematising*, and the manipulation of symbolic expressions (within the *Employ* process of Chap. 1) in the *symbols and formalism* competency. The potential overlap with the *representation* competency was dealt with by removing the reference to representations from the original definition and shifting the focus of this competency to applying procedures, rules and conventions.

Finally, the development of the *reasoning and argument* definition is recorded in Table 4.7. This definition is probably the one that has changed least, other than to give more prominence to the inferential thinking needed to form or evaluate conclusions and arguments. It remains to be seen to what extent the current definition will stand up to use in the context of revisions to the other competencies when the scheme is next tested. It seems likely that further development may be warranted given that this competency has not consistently contributed to the prediction models so far used.

Table 4.7 Development of the *reasoning and argument* definition

Reasoning and argument	
2005	Logically rooted thought processes that explore and connect problem elements to work towards a conclusion, and activities related to justifying, and explaining conclusions; can be part of problem solving process
2007	Logically rooted thought processes that explore and link problem elements so as to make inferences from them, or to check a justification that is given or provide a justification of statements
2013	Drawing inferences by using logically rooted thought processes that explore and connect problem elements to form, scrutinise or justify arguments and conclusions

Level Descriptions

How can different levels of demand for activation of a mathematical competency be identified? This is the second major element of the item analysis scheme, after the general competency definitions. For the scheme to work well it is essential to have a set of level descriptions based on a well-founded and useable set of factors that capture significant aspects of the cognitive requirements of the competencies. They would be factors that do not occur, or that apply at only a low level, with problems for which the competency is less relevant, and that are needed at demonstrably higher levels of intensity in problems where it is more relevant.

In this section, the features used to define the different levels of demand for the six competencies are described. In the following section, Applications of the Scheme, some examples are provided to exemplify application of the scheme to a selection of PISA mathematics problems.

For the *communication* competency, the level of demand for the receptive aspect is described in terms of the complexity of material to be interpreted in understanding the task, the need to link multiple information sources or to move backwards and forwards between information elements (referred to as ‘cycling’). The level of demand for the constructive aspect focuses on the nature and complexity of the parts of the solution process and the explanations or justifications of the result that have to be actively communicated. As with each of the competencies, the descriptions of four levels aim at identifying steps of progression between none or very little of this competency being required, and a substantial requirement for its presence.

In the May 2013 level descriptions (see Appendix 2), the lowest level (level 0) involves understanding short sentences or phrases that give immediate access to the context, where all information is relevant to the task (and no irrelevant information needs to be sifted out) and where the information given is well matched to the task demand. The problem is presented in direct terms that are easily understood and interpreted, without the need for re-reading the text several times in order to understand it, and without the need to forge essential connections between different information elements in the problem statement. At this lowest level the constructive aspect would involve only writing a word, a short phrase or a single number as the problem solution.

The scheme's highest level (level 3) involves understanding more complex text, for example where different information elements need to be understood and linked together in order to proceed, where some of the information may be irrelevant so that a selection and identification process is required, and where logical relationships, for example in the wording of the problem, are more involved. For the constructive aspect, an extended presentation of the solution process may be required, or a coherent explanation or justification of the solution proposed.

The descriptions of levels for the *devising strategies* competency have gone through changes that reflect the substantively changed operational definition to focus on strategic thinking aspect of problem solving and not on problem solving in a more complete sense. Specifically, the wording in the original level descriptions that implies carrying out the strategy devised has been changed. The main challenge here, however, is to identify a plausible gradation of demands. The main variable used to build this gradation is the complexity of the strategy. This has been quantified in terms of the number of identifiably separate stages in the solution process, and complexity is further heightened when those stages themselves involve multiple steps. As part of that complexity, the metacognitive monitoring process needed to keep the solution process on track has also been identified as a contributor to increased demand.

The lowest described level (level 0) for this competency is an example where virtually none of the competency is required. The strategy needed is either stated or obvious from the wording of the problem. The description of level 3 for *devising strategies* refers to a multi-stage strategy that may involve multiple sub-goals. As well as the heightened need for metacognitive control processes at this level, a third indicator of increased demand is in the possible need to evaluate or compare different strategies. These aspects of the description focus on the possibility of a high level of reflection on the problem solving process.

The *mathematising* variable has two separate elements, so the descriptions of graded levels need to pick up both the formulating aspect (transforming features of the context into mathematical form) and the interpreting or validating aspect (discussing the contextual meaning of calculated or given mathematical information). Heightened demand for the formulating aspect is expressed mainly in terms of the degree of guidance provided in the problem statement as to what are the required elements of a mathematical model (assumptions, variables, relationships and constraints). Gradations in the interpreting or validating aspect are arguably less clearly delineated, but the gradation is expressed in terms of the directness of the connection between the mathematical information and the related context, or the degree of creativity required to make that connection. A further element of demand lies in the possible need to evaluate or compare different models, once again implying the need for reflection at higher levels of activation of this competency.

Level 0 for this competency again involves no mathematisation (the situation is purely intra-mathematical, so no translation is required, or the relationship between the mathematical expression of the context and the context itself is not needed to solve the problem). The highest level described envisages the construction of a model where little guidance is given regarding the assumptions, variables, relationships and constraints needed, which must therefore be defined by the problem solver; or validation or evaluation of models in relation to the situation is needed; or there is a need to link or compare different models.

The graded levels of the *representation* competency are based on the complexity of information and interpretation needed in relation to the mathematical representations to be used, the number of different representations that need to be employed and related to each other, and whether there is a need to construct or create an appropriate representation (rather than using given representations) to support the problem solution process.

The lowest described level (level 0) for this competency involves either no use of representations, or very minimal use such as extracting a single numerical value from a familiar table or chart or from text. The level 3 description refers to the need to use multiple representations of complex entities, to compare or evaluate representations (requiring a degree of reflection that can be a feature of higher level demand in a number of competencies), or to create or devise a representation that captures a mathematical entity.

For the *using symbols, operations and formal language* competency, the described levels are based on the degree of mathematical complexity and sophistication of the mathematical content and procedural knowledge required. This competency is clearly subject very much to the educational level of the problem solvers being considered, and the descriptions of levels of activation in the PISA context need to take into account the wide range of levels observed among 15-year-olds in participating countries. Any adaptation of the scheme needs to take the target age range into account for all competencies, but particularly for this one.

The level 0 description is expressed in terms of elementary mathematical facts and definitions, and short arithmetic calculations involving only easily tractable numbers (for example, a requirement to add a small number of one- or two-digit whole numbers) and the use of mathematical rules and procedures that are likely to be very familiar to most 15-year-olds such as the formula for the area of a rectangle. The level 3 description refers to using multi-step formal mathematical procedures that combine a variety of rules, facts, definitions and techniques; and using complex relationships involving variables.

The descriptions of levels of activation of the *reasoning and argument* competency have changed substantially since the initial set of descriptions (in Appendix 1) to reflect the focus of the definition on forming inferences, rather than on general thinking and reasoning steps that might come in to any part of a problem solving process. The levels are described in terms of the nature, number or complexity of elements that need to be drawn together to formulate inferences, and the length and complexity of the chains of inferential reasoning needed.

The description of level 0 envisages inferences of only the most direct kind from given information that lead straight to the required conclusions. The level 3 description requires creating or using linked chains of inferences; checking or justifying inferences; or synthesising and evaluating conclusions and inferences in such a way that draws on and combines multiple elements of complex information. As with other higher-level descriptions this one implies a level of reflection typical of the more demanding levels of activation of a competency.

Application of the Scheme

In this section, a number of PISA problems are presented, an ideal-typical solution process is proposed for each, and a set of ratings for each competency is proposed along with explanation as to why those ratings have been chosen.

M413Q01 Exchange Rate Question 1

The first problem, M413Q01 Exchange rate*,* is shown in Fig. 4.1. This problem originated in the PISA 2003 survey. The problem scenario involves a student preparing to go on exchange from her home country to another country, and needing to change money from one currency to the other. Reading the problem, the two countries are mentioned, together with the names and abbreviations of the two currencies; a conversion rate is given in the form of an equation showing what one unit of the home currency becomes in the other currency; and the question asks how much money the student would get in exchange for 3,000 units of the home currency. To solve the problem, the given model (the exchange rate equation) needs to be used along with some proportional reasoning to scale the 3,000 units up by the amount of the rate. The calculation needed is to multiply 3,000 by 4.2, giving 12,600 ZAR as the required answer. How does the item analysis scheme apply to this problem?

Students need to read and understand the text (e.g. recognising that ‘dollar’ and ‘rand’ are the names of currencies and that SGD and ZAR are their abbreviations, and understanding the link to the equation) and to decide what information is relevant and what is not relevant (e.g. the time period of 3 months is irrelevant to a conversion at the current exchange rate) in order to understand exactly what is required (conversion of 3,000 SGD to ZAR). The material is presented in the order in which it will be used, the text is reasonably straight-forward but with the need to identify and link relevant information, and the solution required is an amount of money. All these features reflect the level 1 description of the *communication* competency. The strategy needed to solve the problem involves using the given equation to scale up the conversion from 1 unit to 3,000 units, which is a straight-forward single-stage strategy that fits the level 1 description of the *devising strategies* competency. To implement that strategy, two main competencies are called in to play. Firstly *mathematising* is needed to transform the given equation into a proportional model enabling the required calculation, and setting up the

Exchange rate

Mei-Ling from Singapore was preparing to go to South Africa for 3 months as an exchange student. She needed to change some Singapore dollars (SGD) into African rand (ZAR).

Question 1

Mei-Ling found out that the exchange rate between Singapore dollars and South African rand was
1 SGD = 4.2 ZAR

Mei-Ling changed 3000 Singapore dollars into South African rand at this exchange rate.
How much money in South African rand did Mei-Ling get?

Fig. 4.1 M413Q01 Exchange rate Question 1 (OECD 2006)

proportional model requires reference to the obvious contextual elements (the two currencies and the money amounts), which fits the level 2 description. Then *using symbols, operations and formal language* is required in order to implement the required calculation (multiplication of a decimal fraction), which fits the level 1 description. The conversion rate equation is a representation of a mathematical relationship, but this aspect of the problem has been taken into account in relation to the other competencies, hence *representation* should be rated at level 0, as should the *reasoning and argument* competency, since the general reasoning needed has been accounted for through the other competencies, and no additional inferences are required.

PM942 Climbing Mount Fuji

Three items from the unit PM942 Climbing Mount Fuji (OECD 2013a) are presented in Figs. 4.2, 4.3, 4.4, and 4.5. This unit was developed for the PISA 2012 survey and used in the main survey. For each problem, a solution process is outlined, and proposed competency ratings are discussed.

The first question, shown in Fig. 4.2, requires calculation of an average number of climbers per day for a given period. To calculate this, the number of climbers is needed (this is given directly for the specified period) along with the number of days (which can be calculated from the dates given). So a strategy would be to find the total number of days, and combine this with the total number of people to calculate (or estimate approximately) the average people/day rate.

Climbing mount fuji

Mount Fuji is a famous dormant volcano in Japan.

Question 1

Mount Fuji is only open to the public for climbing from 1 July to 27 August each year. About 200 000 people climb Mount Fuji during this time.
On average, about how many people climb Mount Fuji each day?

A 340
B 710
C 3400
D 7100
E 7400

Fig 4.2 PM942Q01 Climbing Mount Fuji Question 1 (OECD 2013a)

Climbing Mount Fuji

Question 2

The Gotemba walking trail up Mount Fuji is about 9 kilometres (km) long.
Walkers need to return from the 18 km walk by 8 pm.
Toshi estimates that he can walk up the mountain at 1.5 kilometres per hour on average, and down at twice that speed. These speeds take into account meal breaks and rest times.
Using Toshi's estimated speeds, what is the latest time he can begin his walk so that he can return by 8 pm?

Fig. 4.3 PM942Q02 Climbing Mount Fuji Question 2 (OECD 2013a)

From 1 July to 27 August we have all of July and 27 days of August. There are 31 July days (this is real-world knowledge that must be brought to the problem), and the 27 August days (inferred directly from information given). The total (from an arithmetic calculation adding two two-digit numbers) is 58.

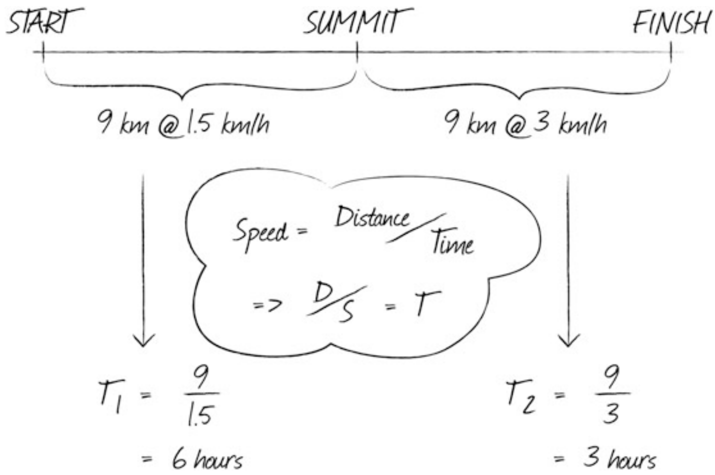


Fig. 4.4 Calculation process for PM942Q02 Climbing Mount Fuji, Question 2

Climbing Mount Fuji

Question 3

Toshi wore a pedometer to count his steps on his walk along the Gotemba trail.
His pedometer showed that he walked 22 500 steps on the way up.
Estimate Toshi's average step length for his walk up the 9 km Gotemba trail. Give your answer in centimetres (cm).

Answer ________________ cm

Fig. 4.5 PM942Q03 Climbing Mount Fuji Question 3 (OECD 2013a)

A model is needed to express the people per day rate, which can be established by applying a fairly obvious definition to express the rate as the number of people divided by the number of days (for the same period). This is a simple model with constraints and variables given directly in the question. The model can be written as average rate = number of people/number of days.

For the rate calculation, one suitable approach might be to calculate 200,000/60 (=3,333), and 200,000/50 (=4,000), and see if only one of the options lies between them. Three thousand four hundred is the only option that lies between these, so response C should be selected.

For *communication,* contextual information must be read and some can safely be ignored (for example, Mount Fuji's name, its dormancy, and its location). The task objective is expressed simply and clearly, with the need only to recognise the two critical bits of information (how long it is open, and the number of visitors in that time), and to form an 'average'. No expressive communication is required. The communication level required is more than zero because of the extraneous information, and the two elements to be combined, but definitely not higher—no cycling through the material is needed because of the simple and straight-forward presentation of information. Level 1 is proposed.

A strategy is needed, involving two distinct but straightforward steps: find the number of days, and combine that with the number of people to form a rate (and then compare these with the given response options). The strategy is not explicitly stated in the text, but not far from obvious; nevertheless it does involve two steps, and hence fits the level 2 description.

A model must be constructed for the rate, but the variables to use seem obvious (people, days), indeed for some students this would constitute a definition; and it also seems a small assumption to suppose that the average number of people per day can reasonably be estimated from the stated period in which Mount Fuji is open to the public. The question can safely be interpreted to mean 'average people per day for the open period' rather than to consider applying the average across the rest of the year when no people would be visiting. This fits level 1 for *mathematising*.

Reading numbers directly from the text is similar to reading isolated values from a graph or table. No transformation into any other specific form is needed, other than to do what is required for the modelling step and for the calculation. The level 0 description for *representation* fits well.

For the *using symbols, operations and formal language* competency, some level 0 calculation is needed (adding two 2-digit numbers), along with some external knowledge (the number of days in July), and the division calculation might lead to decimal results. On balance, level 1 seems to apply.

The *reasoning and argument* competency is proposed at level 0, since the general reasoning steps involved in the assumption about the period for which the calculation is needed are taken as part of *mathematising*, and other general reasoning steps are taken as part of the strategic thinking and the calculations. No additional reasoning steps are involved.

PM942Q02 Climbing Mount Fuji, Question 2

The second question from this unit, shown in Fig. 4.3, involves planning a climb up the walking trail and back to ensure returning by a specified time. The walk comprises two components each 9 km in length, but they are traversed at different speeds hence taking different times. A constraint given in the stimulus is the 8 pm 'latest return time'.

Information is also given about the speed for the two segments of the walk, one being double the other, and that information should be useful for calculating the time it will take, and counting back by that amount from the 8 pm limit. This strategy should be effective. The phrase 'latest time' can be interpreted to mean the time without any rests (other than meal breaks and rest times mentioned in the question as already included in the given average speeds) and with no stopping at the top.

A small sketch such as that shown in Fig. 4.4 helps to represent the given information and to transform the context information (distance at specified speed) to a mathematical form that will provide a way to calculate the time taken (S = D/T). It is helpful to rearrange that formula to D/S = T, in order to calculate the time taken in each of the two components of the walk. This is done for each component, and the two results combined give a total time of 9 h. Finally, the total 9 h needs to be 'subtracted' from the end time. Nine hours before 8:00 pm is 11:00 am, giving a 'latest departure' of 11:00 am.

Some cycling among text elements is needed to understand the task—it contains multiple elements that need to be linked (the distances given, the time constraint, the speed information, and the objective of the question). No expressive demand is made beyond presenting a simple numeric answer. Level 2 seems appropriate for *communication.*

For *devising a strategy,* the solution strategy is somewhat involved, since it has two separate stages: using the given distance and speed data for each segment of the walk to calculate the total walking time, then putting this with the timing constraint to calculate a start time. This is more than the level one description, fits the level 2 description quite well, but probably does not yet amount to the complex multi-stage strategy envisaged for level 3.

Two distinct modelling steps occur here. The first is in formulating the distance/speed/time relationship mathematically (here the constraints are clear, and the variables are spelled out fairly directly); and the second is in translating the calculated distance into a 'latest departure time', which uses reference to the latest finish time and an assumption like 'no more breaks or rests' in order to implement the 'latest time' condition. Having both of these modelling steps leads to level 2 for *mathematising*, but each of them separately might constitute only level 1.

Even though the solution process described includes construction of a simple representation of the given information to help understand and think through the relationships, this was not required and therefore should not be counted as part of the item demand. In this case, level 0 is appropriate for *representation.*

For *using symbols, operations and formal language* level 2 is proposed. The solution process outlined involved writing down, then manipulating the formula connecting D, S, T, and substituting into it (twice); then performing a reasonably simple time subtraction. 'Employ multiple rules, definitions, etc. (including repeated application of lower level calculation)' seems to fit better than 'apply multi-step formal procedures combining a variety of rules, facts, etc.' Each of these by itself might fit level 1, but the requirement for the repeated substitution and the time calculation takes it beyond level 1.

General reasoning steps are needed to formulate a strategy, establish the models needed, and to carry out the calculations required, but no further inferences are needed, so the *reasoning and argument* competency is proposed as level 0.

PM942Q03 Climbing Mount Fuji, Question 3

In the third question of this set, data from a pedometer are given showing the total number of steps taken by a walker, and the question asks for an estimate of that walker's average step length. Given that it takes 22,500 steps to walk the total distance, the average step length will be the distance divided by 22,500. The calculation can be completed by converting 9 km into centimetres (it is 900,000) and dividing this by 22,500, the number of steps, giving an average step length of 40 cm.

The *communication* competency is proposed as level 1 because of the need to link the separate elements in the question statement. The receptive aspect involves recognising one sentence as providing contextual information that is not relevant to answering the question, and bringing together information in three other sentences in order to know what is needed, including the instruction about units. The constructive aspect involves presenting a simple numerical result.

The strategy to calculate an average step-length is a single stage strategy to combine the given elements (divide the total distance by the number of steps). This strategy is not explicitly given, but it seems straight forward. The *devising strategies* competency is therefore proposed as level 1.

The situation model described earlier (distance walked = total distance covered by 22,500 steps) leads directly to the required mathematical model (average step length = distance divided by number of steps), which uses only given variables, and the required relationship seems obvious. This leads to level 1 for *mathematising*.

No additional *representations* are given or required other than extraction of data for the model and for the calculation, so level 0 is appropriate.

Level 2 is proposed for *using symbols, operations and formal language,* which involves a division with large numbers, either after a conversion of units or followed by such a conversion to ensure that the required units are obtained (this step involves drawing on relevant knowledge), and this constitutes 'using multiple rules, definitions, results . . . together' so level 2 rather than level 1 (as each of these calculation steps would be by itself).

A small inference is made using reasoning about one aspect of the problem by using two mathematical entities (count, distance in the required units) to calculate the value required (length per step). This fits with the level 1 description for *reasoning and argument*. Other general reasoning steps (particularly to support the unit conversion) have been taken into account in the previous competency.

Future Steps

The scheme as currently described is presented in Appendix 2. It includes introductory text explaining for each variable some broad features, specific advice about what is and what is not included in the scope of the variable, a summary of the features that drive change across the levels of the variable, a variable definition, and level descriptions.

For these revisions to be tested, a set of annotated items now must be developed to exemplify each competency definition and the assignment of levels for each competency, to guide future uses of the scheme. Some examples have been provided in the previous section of this chapter. While it might be expected that any future use of the scheme would generate results at least as good as those produced in the applications of the previous versions, that expectation must now be tested empirically. Results of that analysis should inform the research team as to the directions needed to further develop and improve the scheme and its documentation. A description of how the scheme is most effectively applied is also needed, since it seems likely that different application methods can lead to different rating outcomes.

A number of wider developments should also be considered. Some action has been taken by independent research teams to apply the scheme, and the results of such independent use will certainly be informative in planning further documentation and development of the scheme. Wider use of the scheme would be very beneficial.

Further research into the drivers of demand within each of the competencies would also be highly beneficial. For example, the elements that make up the descriptions of the four levels of activation of each competency may not yet focus on the most important variables underpinning gradations in competency demand.

It is an open question as to whether the scheme could be used to analyse the mathematical demand of items other than PISA-like items, and items designed for use by students at a different age. The kinds of modification needed for other applications such as these warrants investigation.

Of course other potential uses of the scheme might be the subject of future research. It has already been shown that the use of the scheme can help to improve the targeting of test development procedures, and can improve the efficiency and effectiveness of test development processes (see Chap. 7 by Tout and Spithill in this volume). A similar kind of use could be made by test developers and by teachers in devising assessment items, to check that the items meet criteria related to difficulty and that they elicit mathematical behaviours related to each of the competencies.

One potential importance of the results described in this chapter and in other reports of the analysis of ratings generated from the scheme as it has developed, lies in the implications for mathematics classroom teaching and learning practice. It seems clear that the six competencies described here are very strongly related to the cognitive action taking place as students attempt to solve mathematics problems. It seems obvious, particularly if this finding is reproduced by other researchers, that these competencies should legitimately be taking a prominent place in mathematics teaching and learning, and efforts should be directed to the conscious and visible development of these mathematical competencies among our students. Emphasis is already given to teaching the elements of the *symbols and formalism* competency and perhaps also the *representation* competency. Teaching and practising *mathematisation* requires extensive use of real-world problems, which happens in some but not all mathematics classrooms. Opportunities for practising the *communication* competency, as well as *devising strategies* and *reasoning and argument* are perhaps less commonly observed. An important challenge for the future will be to ensure that teachers teach and provide practice opportunities for each of these competencies, as a way of building levels of mathematical literacy in our students.

Appendix 1: Initial Competency Definitions and Level Descriptions (April 2005)

Reasoning and Argumentation: Logically rooted thought processes that explore and connect problem elements to work towards a conclusion, and activities related to justifying, and explaining conclusions; can be part of problem solving process
0: Understand direct instructions and take the actions implied 1: Employ a brief mental dialogue to process information, for example to link separate components present in the problem, or to use straightforward reasoning within one aspect of the problem 2: Employ an extended mental dialogue (for example to connect several variables) to follow or create sequential arguments; interpret and reason from different information sources 3: Evaluate, use or create chains of reasoning to support conclusions or to make generalisations, drawing on and combining multiple elements of information in a sustained and directed way
Communication: Decoding and interpreting stimulus, question, task; expressing conclusions
0: Understand short sentences or phrases containing single familiar ideas that give immediate access to the context, where it is clear what information is relevant, and where the order of information matches the required steps of thought 1: Identify and extract relevant information, and use links or connections within the text, that are needed to understand the context, or cycle between the text and other related representation/s; some reordering of ideas may be required 2: Use repeated cycling to understand instructions and decode the elements of the context; interpret conditional statements or instructions containing diverse elements; actively communicate a constructed explanation 3: Create an economical, clear, coherent and complete presentation of words selected to explain or describe a solution, process or argument; interpret complex logical relations involving multiple ideas and connections
Modelling: Mathematising, interpreting, validating
0: Either the situation is purely intra-mathematical, or the relationship between the real situation and the model is not needed in solving the problem 1: Interpret and infer directly from a given model; translate directly from a situation into mathematics (for example, structure and conceptualise the situation in a relevant way, identify and select relevant variables) 2: Modify or use a given model to satisfy changed conditions; or choose a familiar model within limited and clearly articulated constraints; or create a model where the required variables, relationships and constraints are explicit and clear 3: Create a model in a situation where the assumptions, variables, relationships and constraints are to be identified or defined, and check that the model satisfies the requirements of the task; evaluate or compare models

(continued)

Problem solving: The planning, or strategic controlling, and implementation of mathematical solution processes, largely within the mathematical world
0: Direct and obvious actions are required, with no strategic planning needed (that is, the strategy needed is stated or obvious) 1: Identify or select an appropriate strategy by selecting and combining the given relevant information to reach a conclusion 2: Construct or invent a strategy to transform given information to reach a conclusion; identify relevant information and transform it appropriately 3: Create an elaborated strategy to find an exhaustive solution or a generalised conclusion
Representation: Concrete expression of an abstract idea, object or action; a transformation or mapping from one form to another; can be part of modelling or PS
0: Handle direct information, for example translating directly from text to numbers, where minimal interpretation is required 1: Make direct use of one standard or familiar representation (equation, graph, table, diagram) linking the situation and its representation 2: Understand and interpret or manipulate a representation; or switch between and use two different representations 3: Understand and use an unfamiliar representation that requires substantial decoding and interpretation, or where the mental imagery required goes substantially beyond what is stated
Symbols and Formalism: Activating and using particular forms of representation governed by special rules (e.g. mathematical conventions)
0: No mathematical rules or symbolic expressions need to be activated beyond fundamental arithmetic calculations, operating with small or easily tractable numbers 1: Make direct use of a <u>simple</u> functional relationship (implicit or explicit); use formal mathematical symbols (for example, by direct substitution) or activate and directly use a formal mathematical definition, convention or symbolic concept 2: Explicit use and manipulation of symbols (for example, by rearranging a formula); activate and use mathematical rules, definitions, conventions, procedures or formulae using a combination of multiple relationships or symbolic concepts 3: Multi-step application of formal mathematical procedures; working flexibly with functional relationships; using both mathematical technique and knowledge to produce results

Appendix 2: Competency Definitions and Level Descriptions (May 2013)

Communication: The communication competency has both 'receptive' and 'constructive' components. The receptive component includes understanding what is being stated and shown related to the mathematical objectives of the task, including the mathematical language used, what information is relevant, and what is the nature of the response requested. The constructive component consists of presenting the response that may include solution steps, description of the reasoning used and justification of the answer provided. In written and computer-based items, receptive communication relates to understanding text and images, still and moving. Text includes verbally presented mathematical expressions and may also be found in mathematical representations (for example titles, labels and legends in graphs and diagrams).

(continued)

Communication does not include knowing how to approach or solve the problem, how to make use of particular information provided, or how to reason about or justify the answer obtained; rather it is the understanding or presenting of relevant information. It also does not apply to extracting or processing mathematical information from representations. In computer-based items, the instructions about navigation and other issues related to the computer environment may add to the general task demand, but is not part of the communication competency.

Demand for the receptive aspect of this competency increases according to the complexity of material to be interpreted in understanding the task; the need to link multiple information sources or to move backwards and forwards (to cycle) between information elements. The constructive aspect increases with the need to provide a detailed written solution or explanation.

Definition: Reading and **interpreting** statements, questions, instructions, tasks, images and objects; **imagining** and **understanding** the situation presented and **making sense** of the information provided including the mathematical terms referred to; **presenting** and **explaining** one's mathematical work or reasoning.

0: Understand short sentences or phrases relating to concepts that give immediate access to the context, where all information is directly relevant to the task, and where the order of information matches the steps of thought required to understand what the task requests. Constructive communication involves only presentation of a single word or numeric result

1: Identify and link relevant elements of the information provided in the text and other related representation/s, where the material presented is more complex or extensive than short sentences and phrases or where some extraneous information may be present. Any constructive communication required is simple, for example it may involve writing a short statement or calculation, or expressing an interval or a range of values

2: Identify and select elements to be linked, where repeated cycling within the material presented is needed to understand the task; or understand multiple elements of the context or task or their links. Any constructive communication involves providing a brief description or explanation, or presenting a sequence of calculation steps

3: Identify, select and understand multiple context or task elements and links between them, involving logically complex relations (such as conditional or nested statements). Any constructive communication would involve presenting argumentation that links multiple elements of the problem or solution

Devising strategies: The focus of this competency is on the strategic aspects of mathematical problem solving: selecting, constructing or activating a solution strategy and monitoring and controlling the implementation of the processes involved. 'Strategy' is used to mean a set of stages that together form the overall plan needed to solve the problem. Each stage comprises a sub-goal and related steps. For example a plan to gather data, to transform them and to represent them in a different way would normally constitute three separate stages.

The knowledge, technical procedures, mathematising and reasoning needed to actually carry out the solution process are taken to belong to those other competencies.

Demand for this competency increases with the degree of creativity and invention involved in identifying a suitable strategy, with increased complexity of the solution process (for example the number, range and complexity of the stages needed in a strategy), and with the consequential need for greater metacognitive control in the implementation of the strategy towards a solution.

Definition: **Selecting** or **devising** a mathematical strategy to solve a problem as well as **monitoring** and **controlling** implementation of the strategy.

0: Take direct actions, where the solution process needed is explicitly stated or obvious

1: Find a straight-forward strategy (usually of a single stage) to combine or use the given information

2: Devise a straight-forward multi-stage strategy, for example involving a linear sequence of stages, or repeatedly use an identified strategy that requires targeted and controlled processing

(continued)

3: Devise a complex multi-stage strategy, for example that involves bringing together multiple sub-goals or where using the strategy involves substantial monitoring and control of the solution process; or evaluate or compare strategies

Mathematising: The focus of this competency is on those aspects of the modelling cycle that link an extra-mathematical context with some mathematical domain. Accordingly, the mathematising competency has two components. A situation outside mathematics may require translation into a form amenable to mathematical treatment. This includes making simplifying assumptions, identifying variables present in the context and relationships between them, and expressing those variables in a mathematical form. This translation is sometimes referred to as mathematising. Conversely, a mathematical entity or outcome may need to be interpreted in relation to an extra-mathematical situation or context. This includes translating mathematical results in relation to specific elements of the context and validating the adequacy of the solution found with respect to the context. This process is sometimes referred to as de-mathematising.

The intra-mathematical treatment of ensuing issues and problems within the mathematical domain is dealt with under other competencies. Hence, while the mathematising competency deals with representing extra-mathematical contexts by means of mathematical entities, the representation of mathematical entities is dealt with under the representation competency.

Demand for activation of this competency increases with the degree of creativity, insight and knowledge needed to translate between the context elements and the mathematical structures of the problem.

Definition: **Translating** an extra-mathematical situation into a mathematical model, **interpreting** outcomes from using a model in relation to the problem situation, or **validating** the adequacy of the model in relation to the problem situation.

0: Either the situation is purely intra-mathematical, or the relationship between the extra-mathematical situation and the model is not relevant to solving the problem

1: Construct a model where the required assumptions, variables, relationships and constraints are given; or draw conclusions about the situation directly from a given model or from the mathematical results

2: Construct a model where the required assumptions, variables, relationships and constraints can be readily identified; or modify a given model to satisfy changed conditions; or interpret a model or mathematical results where consideration of the problem situation is essential

3: Construct a model in a situation where the assumptions, variables, relationships and constraints need to be defined; or validate or evaluate models in relation to the problem situation; or link or compare different models

Representation: The focus of this competency is on decoding, devising, and manipulating representations of mathematical entities or linking different representations in order to pursue a solution. By 'representation of a mathematical entity' we understand a concrete expression (mapping) of a mathematical concept, object, relationship, process or action. It can be physical, verbal, symbolic, graphical, tabular, diagrammatic or figurative.

Mathematical tasks are often presented in text form, sometimes with graphic material that only helps set the context. Understanding verbal or text instructions and information, photographs and graphics does not generally belong to representation competency—that is part of the communication competency. Similarly, working exclusively with symbolic representations lies within the using symbols, operations and formal language competency. On the other hand, translation between different representations is always part of the representation competency. For example, the act of transforming mathematical information derived from relevant text elements into a non-verbal representation is where representation commences to apply.

While the representation competency deals with representing mathematical entities by means of other entities (mathematical or extra-mathematical), the representation of extra-mathematical contexts by mathematical entities is dealt with under the mathematising competency.

(continued)

Demand for this competency increases with the amount of information to be extracted, with the need to integrate information from multiple representations, and with the need to devise representations rather than to use given representations. Demand also increases with added complexity of the representation or of its decoding, from simple and standard representations requiring minimal decoding (such as a bar chart or Cartesian graph), to complex and less standard representations comprising multiple components and requiring substantial decoding perhaps devised for specialised purposes (such as a population pyramid, or side elevations of a building).

Definition: **Decoding**, **translating** between, and **making use** of given mathematical representations in pursuit of a solution; **selecting** or **devising** representations to capture the situation or to present one's work.

0: Either no representation is involved; or read isolated values from a simple representation, for example from a coordinate system, table or bar chart; or plot such values; or read isolated numeric values directly from text
1: Use a given simple and standard representation to interpret relationships or trends, for example extract data from a table to compare values, or interpret changes over time shown in a graph; or read or plot isolated values within a complex representation; or construct a simple representation
2: Understand and use a complex representation, or construct such a representation where some of the required structure is provided; or translate between and use different simple representations of a mathematical entity, including modifying a representation
3: Understand, use, link or translate between multiple complex representations of mathematical entities; or compare or evaluate representations; or devise a representation that captures a complex mathematical entity

Using symbols, operations and formal language: This competency reflects skill with activating and using mathematical content knowledge, such as mathematical definitions, results (facts), rules, algorithms and procedures, recalling and using symbolic expressions, understanding and manipulating formulae or functional relationships or other algebraic expressions and using the formal rules of operations (e.g. arithmetic calculations or solving equations). This competency also includes working with measurement units and derived quantities such as 'speed' and 'density'.

Developing symbolic formulations of extra-mathematical situations is part of mathematisation. For example, setting up an equation to reflect the key elements of an extra-mathematical situation belongs to mathematisation, whereas solving it is part of the using symbols, operations and formal language competency. Manipulating symbolic expressions belongs to the using symbols, operations and formal language competency even though they are mathematical representations. However, translating between symbolic and other representations belongs to the representation competency.

The term 'variable' is used here to refer to a symbol that stands for an unspecified number or a changing quantity, for example C and r in the formula $C = 2\pi r$.

Demand for this competency increases with the increased complexity and sophistication of the mathematical content and procedural knowledge required.

Definition: Understanding and **implementing** mathematical procedures and language (including symbolic expressions, arithmetic and algebraic operations), using the mathematical **conventions** and **rules** that govern them; **activating** and **using knowledge** of definitions, results, rules and **formal systems**.

0: State and use elementary mathematical facts and definitions; or carry out short arithmetic calculations involving only easily tractable numbers. For example, find the area of a rectangle given the side lengths, or write down the formula for the area of a rectangle
1: Make direct use of a simple mathematical relationship involving variables (for example, substitute into a linear relationship); use arithmetic calculations involving fractions and

(continued)

decimals; use repeated or sustained calculations from level 0; make use of a mathematical definition, fact, or convention, for example use knowledge of the angle sum of a triangle to find a missing angle
2: Use and manipulate expressions involving variables and having multiple components (for example, by algebraically rearranging a formula); employ multiple rules, definitions, results, conventions, procedures or formulae together; use repeated or sustained calculations from level 1
3: Apply multi-step formal mathematical procedures combining a variety of rules, facts, definitions and techniques; work flexibly with complex relationships involving variables, for example use insight to decide which form of algebraic expression would be better for a particular purpose

Reasoning and argument: This competency relates to drawing valid inferences based on the internal mental processing of mathematical information needed to obtain well-founded results, and to assembling those inferences to justify or, more rigorously, prove a result.

Other forms of mental processing and reflection involved in undertaking tasks underpin each of the other competencies. For example the thinking needed to choose or devise an approach to solving a problem is dealt with under the devising strategies competency, and the thinking involved in transforming contextual elements into a mathematical form is accounted for in the mathematising competency.

The nature, number or complexity of elements that need to be brought to bear in making inferences, and the length and complexity of the chain of inferences needed would be important contributors to increased demand for this competency.

Definition: **Drawing inferences** by using logically rooted thought processes that explore and connect problem elements to **form, scrutinise** or **justify arguments** and conclusions

0: Draw direct inferences from the information and instructions given
1: Draw inferences from reasoning steps within one aspect of the problem that involves simple mathematical entities
2: Draw inferences by joining pieces of information from separate aspects of the problem or concerning complex entities within the problem; or make a chain of inferences to follow or create a multi-step argument
3: Use or create linked chains of inferences; or check or justify complex inferences; or synthesise and evaluate conclusions and inferences, drawing on and combining multiple elements of complex information, in a sustained and directed way

References

Blum, W., & De Lange, J. (2003). *Item difficulty in PISA mathematics*. Unpublished meeting document of the PISA Mathematics Expert Group (October).

Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2013). Chapter 7: Task analysis in COACTIV: Examining the potential for cognitive activation in German mathematics classrooms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (Mathematics teacher education, Vol. 8). New York: Springer.

Organisation for Economic Co-operation and Development (OECD). (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2006). *PISA released items*. http://www.oecd.org/pisa/38709418.pdf. Accessed 3 Dec 2013.

Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 released mathematics items*. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf. Accessed 8 Oct 2013.

Organisation for Economic Co-operation and Development (OECD). (2013b). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Turner, R. (2012). *Some drivers of test item difficulty in mathematics*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Vancouver, 13–17 Apr 2012. http://works.bepress.com/ross_turner/17. Accessed 23 Nov 2013.

Turner, R., & Adams, R. J. (2012). *Some drivers of test item difficulty in mathematics: An analysis of the competency rubric*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Vancouver, 13–17 Apr 2012. http://research.acer.edu.au/pisa/7/. Accessed 23 Nov 2013.

Turner, R., Dossey, J., Blum, W., & Niss, M. (2013). Using mathematical competencies to predict item difficulty in PISA. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on PISA: Research outcomes of the PISA research conference 2009* (pp. 23–37). Dordrecht: Springer.

Chapter 5
A Research Mathematician's View on Mathematical Literacy

Zbigniew Marciniak

Abstract This chapter provides a personal account of how the views of a pure mathematician on good mathematics education for all students changed through experiences with PISA. Marciniak describes those elements of his own mathematics education that attracted him to mathematics and his own disregard for applications to the real world. His close experience of how students perform on PISA problems have highlighted the difference between significant mathematics and complicated mathematics, and the weakness of educational systems that use a 'catch the fox' paradigm designed primarily for the most talented. It is not true that students who can solve advanced problems can necessarily solve problems that appear simple when analysed only from the point of view of the required mathematical tools. Marciniak has changed his view so that he now sees the ability to employ mathematics when necessary to be the crucial aim of mathematics education for all.

The Charm of Mathematics

As is probably typical for professional mathematicians, mathematics has occupied most of my adult life. It charmed me with its unique beauty in my youth and has kept me under its spell ever since. People outside mathematics usually do not realise that working in pure mathematics has a lot to do with emotions. We usually pick our problems guided solely by curiosity and their aesthetic beauty. However, the 'queen of sciences' likes to be misleading: ideas elude us for a long time as splendid concepts and then most of them end up as misconceptions stemming from a well-hidden error. Nevertheless, once in a while, we are lucky: the idea is right and we get a solution that has previously escaped the efforts of our colleagues. The strike of adrenaline on such, unfortunately rare, occasions is the best reward for the earlier struggles.

Z. Marciniak (✉)
Instytut Matematyki, Uniwersytet Warszawski, Banacha 2, 02-097, Warsaw, Poland
e-mail: zbimar@mimuw.edu.pl

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_5

This kind of intimate cohabitation with mathematics is probably partly responsible for the fact that research mathematicians, especially those in pure mathematics, have quite a strong view of what they mean by mathematics and that makes them (us) quite reluctant to negotiate that view. As being a mathematician includes a permanent self-education process, we perceive mathematics as a path, along which we started our travel some years ago. In consequence, we cannot clearly distinguish our initial mathematical education from our further self-development process as a conscious researcher. One thing, however, is clear to us: our mathematics education was the right one! The proof is that it got us where we are today.

It is quite easy to describe those elements of my mathematics education that 'tasted' the best and which probably made me a mathematician. First of all, those experiences offered me the beauty of mathematics, its clarity and precision. Secondly, as opposed to many other school subjects, mathematics did not refer to any other authority. I was able to judge the truth by myself, according to very simple and clear rules. Next, I was challenged with very nice problems for which I did not know the answer, but which would eventually give in to the pressure of my thinking. Success here makes you feel that there are no obstacles that cannot be eventually overcome. In all of this work, for me, the critical feature was the beauty of the problem and the surprise that it was hiding; the realistic context of the problems was something that I did not care about.

What Is Good Mathematics Education?

I do still believe that an education like the one I received is the optimal education path for a future pure mathematician. However, many of my pure mathematician colleagues very strongly believe that this is the universal prescription for good mathematics education for everyone. I have to admit that for many years my own opinions on good mathematical education were similar. However, my encounter with PISA has changed my opinions on that matter; it stimulated my thinking on the subject and I came to the conclusions presented below.

The first remark is quite simple. In large part, my strong convictions about mathematics are based on the appreciation of the beauty of mathematics. Can we, however, expect that every student will share this view, even if we make the (completely unrealistic) assumption that every teacher is able to present this beautiful science as such? No reasonable person would expect that each student will become a great fan of Bach or Picasso; the same must be true about mathematics. Thus founding the teaching of mathematics on the aesthetic fascination must, in general, end in failure.

I remember getting my first contact with PISA. That was in Lisbon, in May 2002, where the items prepared for the 2003 assessment were presented. (By the way, it is a good tradition of PISA to include mathematicians from the participating countries in the large teams judging items.) Of course I knew there was a document called the "Mathematics Framework", but I was then convinced that the items will tell me all by themselves. To my surprise, and then disappointment, the items seemed to me just trivial. They were not like the problems that I valued most—with purely mathematical context, requiring a smart application of mathematics appropriate to 15-year-olds. I shared this view with one of my colleagues from another country. His reaction was very intelligent; he said: "I know what you mean. But, are you sure that the students to whom you offer your problems would have no difficulties with the items you consider trivial?" I knew right away that the answer must be negative. I had seen the outcomes of the PISA 2000 assessment in Poland. Students, who according to the school curriculum were expected to deal with reasonable facility with complex problems about 'speeds of trains going from city A to city B', were not able to correctly do simple percentages.

The above exchange touches on the fundamental problem of mathematics education policy. Some people prefer what is referred to in Polish idiom as the 'catch the fox' paradigm: you set up the school program so that the most talented profit most; the others just strive to get as much as they can. The talented students are the leading hounds or maybe the fox—all the teacher's attention is focused out there at the front. The other students are the big pack of hounds, running along behind the main action and keeping up as best they can. In the 'catch the fox' paradigm, it is assumed that talent is what matters and that others will fail on many occasions, because their mental capacities simply do not allow them to master solving sophisticated problems in the regular instruction time. They can get a passing mark because a student needs to master only a certain part of the program for that. This paradigm can be very comfortable for some program makers; anything will fit, because the sky is the limit for the possibilities of the best. This paradigm was present for decades in the Polish national curricula, especially in secondary schools. A similar approach was also evident in other Central European countries.

This system was also based on the assumption that a student who was asked to solve sophisticated problems will be able to solve easy ones without much trouble. PISA is a cold shower for those who believe that! Mathematics programs in most countries offer at some point quite advanced mathematical procedures, like for example investigating the variation of a function by studying its first and second derivative. However, the very same students who can learn to do this may have difficulties with calculating a given percentage of a number—a skill much more often encountered in practical life. Even if it were the only advantage of the PISA assessment, it would still be worthwhile participating, for the single reason to learn this lesson.

Today, in European and many other countries, the idea of qualification frameworks has become the main interface between education or training and the labour market. In this setting, the language of learning outcomes becomes crucial. In other words, the attention is shifting from the education process to its results. It is not so important what we are trying to teach students and how complex is the mathematics that is intended to be taught. Much more important is what the students are effectively able to learn.

Mathematical Literacy as the Learning Outcome for All Students

Then, sooner or later, one must ask the question about the purpose of teaching mathematics to all students and about the expected learning outcomes of this process. Clearly, only a very small fraction of them will become mathematicians and even fewer pure mathematicians. What should all the others students learn? From this perspective, I personally find the idea of mathematical literacy to be a brilliant answer. It offers the following perspective on mathematical skills: they are only worth as much as you are able to employ them when needed in your life.

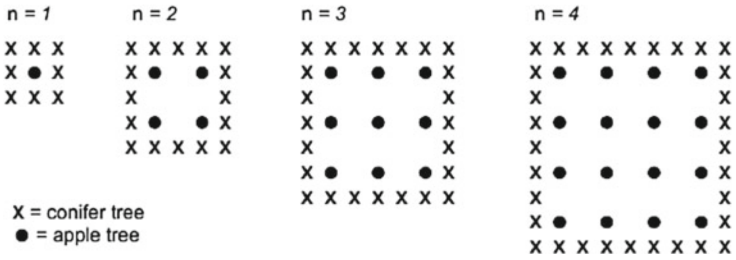
It should be stressed right away, that this mathematical literacy point of view defines no glass ceiling for the skills. Some people seem to think that all you need from mathematics to deal with real life are the basic arithmetic operations with percentage calculations at the most ambitious end of the list, and maybe the measurement properties of the basic geometric figures. The term ‘literacy’ might suggest that absolutely minimal skills are meant, as opposed to ‘illiterate’ which means the lack of the most basic skills. As emphasised in Chap. 1 of this volume, this is not the definition of the PISA Mathematics Framework (OECD 2013).

The PISA set of items shows how wrong those people are. PISA items cover a very wide range of authentic situations, in which you have to invoke mathematical thinking or operations in order to be successful. This process is nicely described in the PISA 2012 Mathematics Framework (OECD 2013) in terms of the modelling cycle. Many of the PISA items require the students to invoke mathematical reasoning or strategic thinking to solve them. In fact, after spending over 10 years working on PISA items, as a member of the Mathematics Expert Group, I came across many items that required mathematical reasoning and argumentation on a level quite satisfactory even from the perspective of a pure mathematician. As one of many good examples, I indicate unit M136 Apples (OECD 2006) as shown in Fig. 5.1. The problem develops slowly through three steps; solving the last one requires decent mathematical reasoning. This item was used in the main survey for PISA 2000 and then released. The difficulty of the item was 550 score points for Question 1 (above average), 665 score points for Question 2 and 672 score points of Question 3. OECD (2006) gives the coding scheme in full.

I find the ability to employ mathematics when necessary to do so to be the crucial aim of mathematics education. Let us notice that it is valid also at the highest levels of research: we admire our most talented colleagues (in the present and from the past) for their ability to find a surprising connection within our domain and solve a problem by employing a tool or idea that no one thought of trying before. The history of pure mathematics is full of such breakthrough stories.

Let us teach all students to find their breakthrough solutions—of course with all proportions preserved. To achieve this goal we should not rush to fill students’ heads with many dozens of new tools for the possible use in some future. Doing mathematics exercises with new tools is like practising scales on a piano—mastering it gives an artisan’s satisfaction, but rarely an excitement. Pressing for more and

A farmer plants apple trees in a square pattern. In order to protect the apple trees against the wind he plants conifer trees all around the orchard. Here you see a diagram of this situation where you can see the pattern of apple trees and conifer trees for any number (n) of rows of apple trees:



Question 1
Complete the table:

n	Number of apple trees	Number of conifer trees
1	1	8
2	4	
3		
4		
5		

Question 2
There are two formulae you can use to calculate the number of apple trees and the number of conifer trees for the pattern described above:
Number of apple trees = n^2
Number of conifer trees = $8n$
where n is the number of rows of apple trees.

There is a value of n for which the number of apple trees equals the number of conifer trees. Find the value of n and show your method of calculating this.

Question 3
Suppose the farmer wants to make a much larger orchard with many rows of trees. As the farmer makes the orchard bigger, which will increase more quickly: the number of apple trees or the number of conifer trees? Explain how you found your answer.

Fig. 5.1 M136: Apples (OECD 2006, pp. 11–14) (formatting condensed from original)

more tools in school curricula decreases the chance that our students find the taste of mathematics at all. We, people doing research in mathematics, understand the meaning of the taste of success very well.

The idea of evaluating mathematics skills through the ability to use them has many advantages. First of all, it refers to the usefulness of mathematics and provides a proof of such. Many pure mathematicians tend to forget that most of the

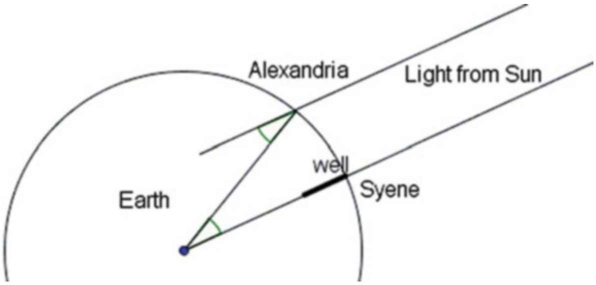


Fig. 5.2 Eratosthenes calculates the circumference of the earth

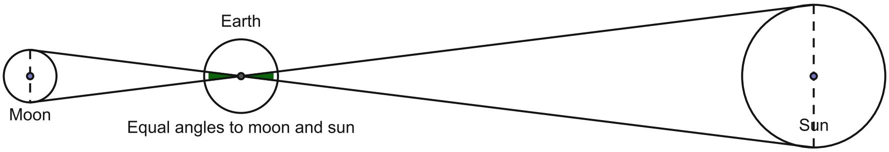


Fig. 5.3 Aristarchus's diagram of moon, earth and sun

most important mathematical ideas were invented or discovered in the process of solving real problems. The theorem about equal angles being created by a line cutting two parallel lines was the basis for Eratosthenes' ingenious calculation of the Earth's circumference around 240 BC (see Fig. 5.2). He knew about a deep well in Syene where the sun only shone on the bottom at the solstice, and he knew the distance of Syene from Alexandria, and could observe the angle of the sun there at the critical moment. This information enabled him to get a very good estimate for the circumference of the Earth.

The theorem of Thales on proportional segments on the arms of an angle cut out by parallel lines was the main mathematical tool used by Aristarchus of Samos in 300 BC to estimate the distances of the Earth to the Sun and Moon. From his discovery that the Sun was much larger and further away, he concluded that probably it is the Earth that rotates around the Sun and not the other way around. Because the angles are equal in Fig. 5.3 (known from solar eclipses), Thales' theorem says that the ratio of the distance of the sun from the earth to the distance of the moon from the earth is equal to the ratios of the sun diameter to the moon diameter. He combined this information with other observations from eclipses and the phases of the moon to draw his conclusion (Protasov 2010). How many teachers of mathematics have ever heard of that? How many research mathematicians remember that? Also many more fundamental examples can be offered from modern times: the idea of a general smooth manifold and its geometry was developed by Riemann and Poincare to satisfy the needs of advanced mechanics and cosmology. Today is no different. The fast developing non-commutative geometry, building the ideas corresponding to measure, topology, distance and differential geometry in the context of non-commutative algebras is just a response to the needs of quantum physics.

I see deep sense in showing our students the relation of mathematics to real life. In fact, many of them may be surprised how often they successfully invoke their mathematical skills. In fact, any reasoning of the type "if. . .then. . ." or "it is not so, because. . ." has some mathematics content.

The types of pure mathematics problems that I mentioned at the beginning of this paper as enjoying so much, are artificially (and skilfully) composed like a chess problem: you are given a position and perhaps must invent how to make the checkmate in two moves. Whatever we say, it is just an intellectual game. Some people, like me, find deep satisfaction in playing such games. All others like to see a purpose.

Final Thoughts

Finally, I want to make two more points. The first is the following. Over the long period of work in the PISA Mathematics Expert Group I have learned to appreciate the really hard research in mathematics education. Compared to the problems those people try to solve, my non-commutative algebra problems look like child's toys: clearly formulated, simply stated with the only catch being that no one knows how to solve them. The mathematics educators' problems have completely different nature: the basic difficulty is probably to identify the problem. Even then, there are many 'ifs' and 'buts', because it inevitably touches on diverse areas including neuroscience, psychology or even sociology. And then a solution to such a problem will have exceptions (counterexamples!) and yet it still has a value. I learned a great lesson in this area. I was most fortunate to learn from the best of the best: Mogens Niss, Werner Blum, Kaye Stacey, Ross Turner, Sol Garfunkel, Bill Smith—to name a few of them. Thank you, friends!

And the last comment. Once we, the Mathematics Expert Group members, accepted the mathematical literacy perspective, we got an unexpected bonus: during more than 10 years of meetings I do not remember even a single discussion concerning the differences of the content of mathematics curricula in different countries. After all, what counts in PISA is the ability of a student to resolve a situation, whichever mathematics tools stand at his/her disposal. The problems we decided to offer to test the students were nearly always of the kind that several different successful approaches were possible and could be effectively used by students. By the way, that shows great flexibility of mathematics, even at the elementary level. That is one more reason that I am proud of being a mathematician.

References

Organisation for Economic Co-operation and Development (OECD). (2006). PISA released items—mathematics. http://www.oecd.org/pisa/38709418.pdf. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD). (2013). 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. Paris: PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en. Accessed 23 Aug 2013.

Protasov, V. (2010). Geometria zviozdnogo nieba [Geometry of the starry sky]. *Kvant, 2010*(2), 14–22. http://kvant.mccme.ru/pdf/2010/2010–02.pdf. Accessed 24 Sept 2013.

Part II
Implementing the PISA Survey: Collaboration, Quality and Complexity

Introduction to Part II

In this part aspects of the implementation of the PISA survey from various insider perspectives are presented. The contributions provide insight from the perspective of a leading test development agency and its collaborators, into the mathematics test development for the world's largest international educational survey. The main themes that emerge from this part are collaboration, complexity and quality. Even just in its mathematics component, through the stages of framework to items to survey to data, the PISA survey requires collaboration of people from many different countries and with many different skill sets. The contributions in this part highlight the many different types of expertise required, as well as the way in which collaboration and critique from around the world is brought in to optimise the validity and relevance of the end product. The complexity of the undertaking is both in the theoretical understanding of the task and also in the logistics of delivery that arise, and several of the chapters describe some of these aspects. A raft of quality assurance measures are employed so that the product of all this effort will provide sound data for educational decision making. A further overarching message of this part is that participation in the PISA survey provides an array of training and development opportunities to individuals, organisations and systems. Later Part III of this volume details how some of this learning is also now applied within national assessments. In addition, PISA has provided an impetus to conceptual and technical innovation and invention at various points of the survey process.

Ross Turner has led the development and delivery of the mathematics component of PISA for the lead international contractors since 2000 not long after PISA's inception, and in the lead-up to the first PISA administration. He opens the part in Chap. 6 with an overview of the different activities that various participants and contributors engage in to bring the PISA mathematics survey to fruition. Some of the mystery surrounding the development and implementation of this high profile endeavour is exposed, and some commonly expressed concerns about PISA are answered. In Chap. 7, item developers Dave Tout and Jim Spithill use the stories of two PISA items to provide insights into the intricate process of developing PISA mathematics test items, both paper-based and computer-based. Those ideas are picked up and developed in greater depth in Chap. 8 by Caroline Bardini who provides additional examples, analysis and commentary related to the computer-based assessment which was a new innovation for PISA 2012. In Chap. 9, Agnieszka Sułowska offers an insightful overview, from the perspective of an experienced national head of coding for four PISA surveys, of the processes and issues involved in transforming student responses to the PISA mathematics items into data for analysis. She explains how different this coding process is to the process of a teacher marking student work. The chapter also illustrates the investigation into students' mathematical thinking that is sometimes required for accurate coding. In addition to practical advice for coding of large scale assessments, and insights into how some PISA items operate in practice, this chapter is significant for its frank discussion of the complexity of the coding task, and the depth of personal experience that it draws on. This part concludes in Chap. 10 with a discussion of one set of the background questionnaire variables introduced into the student questionnaire for PISA 2012. In their chapter, Lee Cogan and Bill Schmidt provide important insights to the *opportunity to learn* sections of the student questionnaire, giving a glimpse of the depth of thought and earlier research that has gone into developing this crucial aspect of the battery of PISA survey instruments. This chapter also highlights the different intentions between the two major series of international mathematics achievement studies, PISA and TIMSS; differences that are reflected in different approaches to opportunity to learn. The curriculum-based TIMSS surveys aim to test content that students are very likely to have had the 'opportunity to learn' while the PISA surveys start with the intention to assess knowledge and skills that are judged most valuable. Information derived from the conjoint analysis of responses to background questionnaire variables and domain (mathematics) variables is what gives PISA much of its power to generate insights leading to policy reform and innovation in teaching and learning.

Chapter 6
From Framework to Survey Data: Inside the PISA Assessment Process

Ross Turner

Abstract This chapter provides an overview of quality assurance mechanisms that have been put into place by the international contractor responsible for implementing the PISA survey. These quality assurance mechanisms aim to ensure the fitness for purpose of the PISA data, derived from over 60 different countries and from students instructed in over 40 different languages in a wide array of schools from education systems that vary quite considerably. The mechanisms reviewed include the frameworks that drive the content of the PISA survey instruments, the processes followed in test item development, student sampling procedures, the mechanisms designed to guarantee comparability of the different versions of test instruments that go into the field in participating countries, steps to ensure test administration procedures are common across all test administration centres, the mechanisms associated the processing and scoring of student responses to the test questions, and processes related to capturing, processing and analysing PISA data. This chapter also brings together into an accessible and consolidated form, information that has been published in a variety of other documents, such as in PISA operational manuals and technical reports. However, the chapter also explains the significance of these processes and the reasons for the decisions, and highlights how they are implemented for mathematics.

Introduction

How is it possible to implement a survey in more than 60 countries that generates measures that are in any sense comparable? This is the challenge faced by the international contractors responsible for implementation of the PISA survey.

This chapter outlines various steps taken by the Australian Council for Educational Research (ACER) and its international collaborators to implement PISA in

R. Turner (✉)
International Surveys, Educational Monitoring and Research, Australian Council for Educational Research (ACER), 19 Prospect Hill Rd,
Camberwell, VIC 3124, Australia
e-mail: Ross.Turner@acer.edu.au

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_6

such a way that meets this challenge based on ACER's experience in delivering the survey across its first five administrations. ACER has been the lead agency in an international consortium awarded the contracts to deliver the first five administrations of the PISA survey (in 2000, 2003, 2006, 2009 and 2012) on behalf of the Organisation for Economic Co-operation and Development (OECD). The author has been a senior manager within the ACER project team from early 2000, so has seen the entire survey administration process from close quarters including two complete survey administration periods in which mathematics was the major survey domain. The story of PISA implementation has several threads. One important thread relates to framework development (see Chap. 1) and the steps of development of the items that end up in the PISA tests, which are discussed in more detail by Dave Tout and Jim Spithill in Chap. 7 of this volume. But the generation of comparable measures in such a large survey program also involves quality assurance procedures related to every aspect of survey delivery: the sampling of survey respondents, the preparation of translated versions of survey instruments, mechanisms to ensure that test administration procedures are the same everywhere, steps to ensure that survey responses are processed and scored using a common set of standards, data capture procedures that ensure the integrity and confidentiality of data generated by the survey, and data analysis methods that guarantee the national and international reports that are generated are fit for purpose. This chapter brings together into an accessible and consolidated form, information that has been published in a variety of other documents, such as in PISA operational manuals and technical reports that are all are available from the PISA website (www.oecd.org/pisa). However, the chapter also explains the significance of these processes and the reasons for the decisions, and highlights how they are implemented for mathematics.

The Starting Point: An Assessment Framework

One essential requirement for a useful measurement of learning outcomes within any domain is clarity about what will be measured. Assessment frameworks have been developed with exactly this need in mind, in order to guide a vast array of local, national and international assessment enterprises. The importance of assessment frameworks is captured by Jago (2009) writing on the history of the frameworks used in the USA's National Assessment for Educational Progress (NAEP) in a piece written for the 20th anniversary of the National Assessment Governing Board.

> [NAEP] frameworks describe the content and skills measured on NAEP assessments as well as the design of the assessment. They provide both the "what" and the "how" for national assessment. Representing the best thinking of thousands of educators, experts, parents, and policymakers, NAEP frameworks describe a broad range of what students learn and the skills they can demonstrate in reading, mathematics, writing, science, history, civics, economics, foreign language, geography, and the arts. (Jago 2009 p. 1)

The PISA Frameworks were initially developed for the first survey administration in 2000, covering the domains of Reading, Mathematics and Science (OECD 1999). The Frameworks have been revised, updated and expanded over the period in which the PISA survey has existed, as the domains have evolved (for example as digital technology has changed the way learning occurs and as computer-based assessment components have been added) and as additional survey domains have been added (for example as a separate problem solving component was added, first in 2003 and again in a computer-based form for PISA 2012). A 'questionnaire framework' has also been developed to provide the rationale for the collection of various elements of background and contextual information gathered through a suite of questionnaires used alongside the cognitive instruments.

The PISA frameworks specify what is to be measured in each of the assessed domains. They define each assessment domain, and from the domain definitions the frameworks spell out in considerable detail the constructs of interest and their key components, the constraints within which those constructs will be understood and approached, and the variables that will be built into the pool of test and questionnaire items developed to generate indicators of the constructs of interest. Each framework provides a blueprint for test or questionnaire development in the relevant test domain.

The Mathematics Framework, as discussed by Stacey and Turner in Chap. 1 of this volume, defines mathematical literacy, the central construct of interest. It outlines mathematical content categories, mathematical processes, and a range of mathematical problem context types, which are taken as constraints within which the assessment of mathematical literacy is to be understood and approached. The Framework also sets constraints related to the range of students to be assessed through PISA (15-year-olds in school) and the consequent span of the mathematical literacy construct that will be targeted, as well as the kinds of mathematical problems that can realistically be used in an assessment of this type. The Framework describes how each of these variables will be arranged and balanced in a PISA survey instrument.

Of course the PISA Frameworks are not handed down as if they were biblically ordained laws. Rather, they are developed through widespread and ongoing consultation processes that involve experts from all participating countries contributing information about the priorities of the assessment domain and the potential basis for an international comparative survey of such a scale. Framework drafts are developed, circulated for comment, and revised, and are only adopted when sufficient buy-in has been achieved to permit the OECD's PISA Governing Board to have confidence that the Frameworks are sufficiently reflective of the interests of all participants.

This consultative and inclusive approach to the development of the PISA Frameworks sets a template of consultation and involvement that is reflected in all aspects of PISA survey development and implementation.

High Quality Survey Instruments

Assessment frameworks provide definition, and must be enacted through the survey instruments developed to generate indicators of the constructs of interest. The processes followed to develop PISA survey instruments further ensure that PISA adheres to the highest standards of technical quality.

As the lead contractor appointed by the OECD to develop and implement the first five administrations of the PISA survey, ACER has led the development of test items in each of the survey domains and the formation of the test instruments so far used in PISA. Typically this has been driven by a team of professional test developers at ACER, working in collaboration with teams of professionals from test development agencies in other locations and other countries, under the guidance of a reference group of international experts. In the case of mathematics, the international Mathematics Expert Group (MEG) includes mathematicians, mathematics educators, and experts in assessment, technology, and education research from a range of countries. Material and ideas for test items in each domain come from a variety of sources: from team members of the various professional test development agencies contracted to contribute, from members of the MEG, and most importantly from teachers and other domain experts in participating countries. All countries that participate in a particular administration of the PISA survey are encouraged to submit items, and many have chosen to do so. These contributions are typically sought through the PISA national centre, which coordinates and manages all PISA-related processes within each participating country. Using material from such a diverse range of sources helps to ensure richness and variety in the pool of material from which a set of test items is built that expresses different approaches and priorities in different countries, as well as different cultural and educational practices, within the orientation and constraints set by the Framework.

The test development teams institute a rigorous process to turn ideas and suggested items into test content. Typically, this starts with a rigorous ‘shredding’ of the draft item by a small panel of developers. That involves scrutinising the material from several angles—clarity of wording, quality of accompanying stimulus material, fit to the Framework, the range of possible responses to the item, and so on. Items are then revised, and sent to one of the other teams to repeat the process. Once draft items showing potential reach that stage, they are then subjected to small-scale field testing with individual students, and with small groups of students through ‘thinking aloud’ methods as described by Rowe (1985) and cognitive interviews. Students of the same age as the intended target for the PISA test are given the draft items, and asked to attempt the items and to ‘think aloud’ as they do this in order to expose their thought processes as they tackle the problem. The test developer records this or takes careful notes for later analysis, and then further probes the students by asking them to articulate further their reaction to particular elements of the problem, the solution method they used, and the steps they took as they attempted the problem. In this way, further insights are gained into the draft item and whether the item is triggering the kind of mathematical thinking and behaviour assumed or sought. Carrying out such a process with students in different countries, and using material presented in different languages, gives significant insight into the merits of the item and its likely usefulness to generate indicators of the constructs of interest. In Chap. 7, Tout and Spithill illustrate this development process with some sample items. A further stage in the development of items uses field testing of sets of items with larger groups of students under test conditions. Such a procedure can be used to trial different forms of an item (for example to test alternative wording). Using several groups of students on whom to field test sets of items is a way to generate useful comparative information about a group of items (for example, the relative difficulty of items) and it can be very helpful in the development or refinement of scoring rubrics that are used to identify the set of possible responses to each item actually observed among responding students.

At the conclusion of this item development process, source versions of the items are prepared in both English and French, and sets of these items are then formed and sent to participating countries for review. The reviews are normally undertaken by national experts in the domain, who are asked to provide detailed feedback on each item including: relevance of the item to the key OECD notion of 'preparedness for life' from the perspective of each participating country; relevance of the item to the local mathematics curriculum; likely interest level of the item to 15-year-old students; the degree of authenticity of the item context from the perspective of the country; whether there are any cultural concerns or other potential sensitivities with the item; whether any translation issues are anticipated; whether any problems are anticipated with the proposed response coding rubric; and a rating of the country's view of the priority for inclusion of the item in the final selection for use in the PISA survey instruments. The information received from these reviews is used to identify items that will be unacceptable to participating countries, and contributes to the selection of items for possible use in the PISA survey.

As a final part of ensuring items of the highest possible quality are available for use in the PISA survey, PISA employs a two-stage process in its implementation of each administration of the survey: a field trial, and a main survey. In the year prior to the main survey, an extensive field trial is administered in every country participating in the survey. A large pool of test items that have successfully gone through the development process described in the previous paragraphs is selected for field testing. Those items are translated into the required national languages, and the translations are verified according to a highly rigorous process, placed in test booklets according to a rotated test design, and administered to several hundred students in each country. Test booklets are formed by putting together four clusters of items, with each cluster representing 30 min of test time, and following a linked rotation design that ensures each cluster appears exactly once in each of the four possible cluster positions, and exactly the same number of times in total, across the set of booklets. Test administration procedures are developed centrally by ACER and its collaborators, and national teams in each country are trained in those procedures to ensure a high degree of consistency across participating countries. Teams of mathematics experts in each country are trained to assign standard codes to each response observed to each item. For more detail of the coding process, see Chap. 9 by Sułowska in this volume. Standardised data capture procedures are implemented, so that ACER receives consistent and reliable data from all test administration centres.

The field trial generates data that can be used for a variety of purposes. Some of those relate to the operational issues involved in test administration within each country, while others relate to the technical qualities of the test material. Extensive analysis of the field trial item responses provides further information on the quality of each country's translation of the test material, and on the psychometric properties of each item. This allows the test developers to understand how the items are likely to actually perform when administered to 15-year-old students: which items generated useful data, what was the empirical difficulty of each item, did any of the items perform differentially with boys compared to girls or with students in one country compared to another (after adjusting for student ability) and so on. Data and information generated from the field trial provide a very strong basis on which to identify the best available items for possible inclusion in the main survey item pool.

After the field trial, a further review of items by experts at the national centre for each participating country is conducted, generating fresh information based on countries' field trial implementation experience relating to any unanticipated translation issues, unexpected difficulties with the coding of student responses to items, or any other problems identified, and providing a new set of priority ratings for inclusion of each item in the main survey item selection.

By the time test items are chosen for inclusion in the main survey instruments, they have been through a development and selection process designed to produce items that are fit for purpose from the perspective of a variety of technical characteristics, their useability for the intended target audience, and their acceptability to relevant stakeholders.

Questions for use in the various background questionnaires are developed using a similar mechanism. Those questions are based on a questionnaire framework that provides a theoretical basis for the background variables of interest, which are used to help understand which students perform at different levels, what characteristics of the students' backgrounds might explain differential performance, and in particular what factors that influence performance might be affected by particular aspects of educational practice and policy settings at the local, regional or national levels. Chapter 10 by Cogan and Schmidt in this volume describes the theoretical foundation and development of the 'opportunity to learn' thread of these questionnaires.

Rigorous Scientific Student Samples

PISA survey instruments (the booklets containing questions about the assessed domains, and the student background questionnaires) are administered to scientifically sampled students in each participating country. Sampling standards are designed by the PISA international contractor to ensure proper coverage of the population of interest, an acceptable level of accuracy and precision in the estimates derived from PISA data, and adequate school and student response rates, and these are applied in each participating country.

The PISA national centre in each participating country obtains a sampling frame that lists all educational settings with students falling inside the age definition for the survey, and provides this to the international PISA contractor that then checks its accuracy. A limited number of school exclusions are permitted for well-defined reasons such as inaccessibility through extreme remoteness, or the existence of political turmoil in a particular part of the country that would make survey administration dangerous or impossible, and any such instances are documented. Steps are taken by the PISA international contractor to verify the accuracy and completeness of each country's sampling frame from independent sources—for example, by comparing the data in the sampling frame with other publicly accessible data.

Sampling for the PISA main survey in each administration then proceeds through two main stages: the international contractor selects a random sample of schools, with the probability of selection being proportional to the number of eligible students; then for each of those sampled schools selects a random sample of eligible students. The number of schools and students required is determined to achieve an acceptable degree of precision in the estimates derived from the survey. Typically about 150 schools are sampled, and within each school a sample of about 35 students are selected, meaning a total of a little over 5,000 students are typically sampled in each participating country. Some countries increase their sample because they are interested in finer-grained information about particular subpopulations, for example several participants take a larger sample in order to get regional or provincial estimates. In Chap. 13 of this volume, Arzarello, Garuti and Ricci describe how such regional information has been used in Italy. Accuracy in the estimates of the location of the measured population (mean score), and the precision of those estimates (the narrowness of the range of possible estimates), are also increased as countries take up the possibility of systematically applying stratification variables to the sampling process, whereby schools are classified and sampled according to variables on which they tend to be similar, such as school type, school size, programme type, school funding source, or location variables. In some cases countries have also used such stratification variables to provide greater detail in their national PISA reports.

A key piece of information captured from every PISA survey administration site is the number of sampled students who actually respond to the survey. Acceptable response rates are defined at both the school and student levels, together with mechanisms for sampling additional schools to substitute for sampled schools that refuse to or cannot participate or where student response rates are unacceptably low. The recorded response rates are used to determine whether the sampling standards have been met in each participating country, and whether the student response rate at each sampled school is sufficient to include data from that school. There have been cases of countries having data excluded from the PISA international reports and database because of failure to meet the response rate standards. For example, data from the Netherlands were excluded from a large number of tables in the PISA 2000 international report, and only included as a separate line in other tables with a footnote indicating “Response rate too low to ensure comparability” (OECD 2001). Several countries have come close to achieving unacceptably low response rates in one or more PISA administrations (including Australia, the Netherlands, UK and USA) and such countries have to work hard to achieve acceptable rates.

In addition, data about the number of respondent schools and students are used to determine sampling weights that are the statistical mechanism applied to PISA data to ensure the sample gives the most accurate possible estimates of the targeted characteristics of the population of interest.

Linguistic Quality Control for Test Materials

A further area requiring explicit steps to guarantee the quality of survey instruments in such a large international survey is the translation and adaptation of test materials in the array of local languages of instruction used in participating countries. In this section, two main aspects of quality assurance related to linguistic quality control in PISA will be discussed. The first is the steps followed as part of the development of test items to take language, cultural and translation issues into account in order to anticipate and minimise potential translation difficulties. Second, the mechanisms used to achieve the highest possible quality across the 85 different national versions of survey instruments (tests and questionnaires) in the 43 different languages that were used in the PISA 2012 survey administration are briefly reviewed.

Linguistic Quality Issues in the Design of Test Questions

PISA test materials may originate in any of a variety of languages and an early step in item development is preparation of an English language version that reflects the intentions of the item’s author and any modifications subsequently introduced by the professional test development teams that work on the item. At an early stage, a parallel French language version is also developed, and these two versions are further adjusted to ensure their equivalence with the help of the advice of content experts fluent in both English and French under the direct guidance of the test developers. These two versions are referred to as the *source* versions of each item. They are tied together using a version control process that means a change to one version causes status changes that ensure the two source versions remain synchronised.

Source versions of each item can be subject to change for a variety of reasons. Some of those relate to specifically mathematical issues inherent in the item. Some relate to the lessons learned about each item as it is used with individuals and groups during the item development process. However another key source of input to items as they develop comes from the accumulated knowledge and wisdom of the linguistic quality control experts engaged by ACER to provide advice on cultural and language factors known to be critical to the design of good test questions that will be used in different languages and different cultural contexts. The objective of this advice is to ensure that the source versions of all test items can be rendered as equivalently as possible into all of the target languages.

The linguistic quality advice given to test development teams covers a range of technical matters, including syntactic issues, vocabulary issues, cultural issues, and even matters related to the presentation of graphics. Guidelines relating to syntax are important because different languages employ different syntactic rules and structures, and this can threaten the equivalence of different language versions. Experience has shown that certain syntactic forms should be avoided wherever possible because the different forms used in different languages make translation extremely difficult. One example is the need to avoid incomplete or hanging stems in a question statement because different languages structure the missing part of an incomplete stem differently. For example in English, the missing part can be at the end, but in other languages such as Turkish the missing part must be at the beginning. This applies particularly commonly to questions presented in multiple-choice format. Another is the problem caused by use of the passive voice. Sentences that contain more than one phrase expressed in the passive voice can be extremely difficult to translate while retaining a comparable level of reading difficulty. Such issues can generally be avoided by transforming the offending phrase or sentence and using only direct wording expressed in the active voice. Another syntactic problem arises from long or unduly complex sentences, since translation can make these even longer and more complex. Often that problem can be solved simply by breaking the long sentence into a number of shorter ones.

The phrasing of questions can also create translation difficulties in particular languages. Questions in English beginning with 'how' (how much . . ., how many . . ., by how many times . . .) can be difficult to translate into some languages. Sometimes this can be resolved by changing to a 'what. . .' question. For example instead of asking 'How much tax is on . . .' it is better to ask 'What is the tax on . . .'. Instead of asking 'How fast does the vehicle go . . .', it is better to ask 'What is the speed of the vehicle. . .'. Questions beginning with 'which' (which of the following . . .) cannot be used because in some languages a word denoting either singular or plural is required, hence giving more information than is contained in the English version. Expressing the source version as 'which one of the following. . .' or 'which one or more of the following. . .' depending on what is intended usually gets around this issue. A similar issue arises according to whether the language requires the noun ending to change or not to change according to whether it is singular or plural.

Vocabulary-related issues can also cause translation problems. For example, common names of plants and animals can be impossible to translate without clear guidance on the object being named (for example, by including the object's Latin name in a translation note). Some technical terms, including mathematical terms, can be difficult to translate where standard usage and definitions may not be in common use in different countries. For example, different types of graphs or charts need to be referred to with care. Similarly, the word ‘average’ can be interpreted differently according to different usage in mathematics classroom in different countries. In some countries ‘average’ would refer to the arithmetic mean, but in others its usage would be a more generic reference to measures of central tendency, perhaps including the median in its interpretation. Words with a technical meaning in English (for example ‘quadrilateral’) are rendered in some languages by words that spell out key features of the definition (‘four sided figure’) so there would be no point in asking a question that including testing whether the student knows the meaning of such a word. Care must also be taken to decide on the implications of using a word that may have an agreed technical definition, but for which common usage may vary (for example, the word ‘weight’ would more correctly be referred to as ‘mass’, but such words may be understood differently according to common usage). Common ways of expressing ranges of numbers (for example whether the boundaries are included in a phrase such as ‘between A and B’) create issues both for the wording of questions, and for the interpretation of responses. The use of metaphors or other ‘figures of speech’ in the wording of questions is another issue that can cause translation problems. For example, the phrase ‘helicopter view’ to denote an overview of a situation without any details may not convey the same meaning in so few words in other languages. Many metaphors tend to be language-specific, and cannot always be translated without making the wording longer or more complex.

A number of issues that might be regarded as *cultural* matters have also been highlighted by the linguistic experts, often related to different levels of familiarity with objects referred to in mathematics problems. For example, a question requiring familiarity with a metropolitan rail system (such as an underground metro) might present very different challenges for students living in a city with such a system compared to students from a remote rural community. The extensive item review processes in which all PISA countries participate tend to pick up those issues. Nevertheless being aware of potential problems of this kind in advance can help to avoid difficulties before they arise.

Finally, even matters related to the preparation of graphic materials need to be considered from a linguistic quality point of view. Graphics typically contain labels and other text, and these must be put together in such a way that they are easily editable by those responsible for translation in each country. Not only that, but the design of the graphic elements in the source versions must take account of language variations such as the maximum length of words or phrases when translated, and the direction in which text is written (left to right, or the reverse as in Arabic). Great care is needed to ensure that graphics are designed to accommodate the different language demands in such a way that the ease of use and interpretation of the graphic is consistent across languages.

Maximising the Linguistic Quality of National Versions

A major quality assurance challenge arises in relation to the need for each country participating in PISA to prepare test materials in the local languages of instruction that can be regarded as comparable to the source versions prepared by the international contractor. Without this, PISA results would have no credibility.

A total of 18 countries involved in the PISA 2012 survey administration used French or English language versions adapted directly from the appropriate source version. Adaptations might include substituting familiar names for people referred to, or changing to the local spelling standard. In all other countries, where other language versions were needed (referred to as the *target* versions), translation experts within each country were responsible for producing a local target language version of each item. Different countries may have used slightly different processes to achieve this. The recommended approach, used in several countries, is to use both the English and French source versions independently to produce two target versions that are then reconciled by an independent translation expert to form a single version. In other countries, two independent translations were generated from one of the source versions (either the French or the English language version), each with cross-checking against the other source version, and the two versions were reconciled by an independent translator into a single version. In several PISA countries that share common languages, these translation tasks were shared by experts in the cooperating countries.

The next stage in the process is to have each reconciled local version verified by an independent expert. This work has been done by one of ACER's consortium partners (cApStAn Linguistic Quality Control) that employs language experts who are all trained in application of the rigorous standards and procedures used in the verification of translated PISA instruments. Personnel fluent in both the target language and at least one, and often both, of the source version languages (English or French) were engaged to undertake the verification, which consists of a detailed comparison of the target and source versions. The team of verifiers met for face-to-face training, and used a specially prepared set of training materials including a common set of guidelines that defined exactly what kinds of things they should look for in evaluating the quality of each translation, and lists of quite specific issues to look for in relation to particular items for which potential translation problems had previously been identified. Verifiers used a common set of categories that defined exactly when an expert judgement was required to approve or reject a proposed translation element, or to seek clarification of the reason for a proposed change, or to refer proposed text to a central authority for further consideration. Categories included such events as text being added that was not in the source version, text missing that should have been there, layout changes, grammar or syntax errors, consistency of word and other usage both within and across units, mistranslations (so that the intended meaning is changed), and so on.

Some countries sharing a common language cooperated in a further procedural variation, whereby the countries using a particular shared language cooperated to prepare a single version that was verified according to the standard processes then adapted (subject to an external approval process) to suit the particular needs of each of the cooperating countries. A similar process also occurred where one country borrowed a verified version from another country having a shared language, and introduced approved adaptations where needed to make it suitable for local use.

Once the verifier interventions had been carefully considered by the national translation experts in each country, and final test booklets were formed from the verified target version, an external final optical check was carried out to ensure that the materials had been correctly assembled into the student booklets, and to identify any remaining errors that had been missed in the national centre.

The translation and verification process described here is undertaken in its fullest and most rigorous form at the stage of preparing materials for the field trial. After the field trial, when a selection from the test material is identified for use in the main survey, a further lighter-touch verification is undertaken that focuses on any changes made to the source versions of the items (or their response coding instructions), and on any errors identified in the local versions as a result of the field trial experience. Because of the complexity of creating comparable items, items that do not perform optimally at the field trial are almost always discarded rather than changed for the main survey.

Common Test Administration Procedures

For an international survey to generate comparable data from the different countries that participate, the procedures through which the survey is administered should be as near as possible to the same in all countries. The PISA survey uses a variety of processes designed to ensure common and high standards of test administration are adhered to everywhere.

Each country that participates in the PISA survey appoints a survey administration team. In some cases this is a team assembled and trained by the PISA national centre, in other cases dedicated test administration agencies might be appointed to administer the survey. The international contractor (ACER and its collaborators) has developed a detailed set of test administration procedures, which are documented in a series of manuals. These are used as the basis for training personnel from each participating country using a ‘train the trainer’ model, so that those responsible for managing test administration in each country are given the same training, and are provided with the same guidelines and instructions. The instructions are explicit and detailed, and include a script that is used in every test administration session in every participating country to introduce the test and get students started on answering the survey questions.

Test administrators can be school personnel, or they can be external staff employed specially to conduct test sessions in a number of schools. In the case where internal school personnel are used as test administrators, guidelines are designed to ensure that teachers do not administer test sessions that contain students they teach in any of the subject areas being tested.

The test administration guidelines cover such matters as protocols for contacting schools to arrange the basic details such as location and time of each test session, and permission from parents for the students' participation in countries where this is required; packaging, transport, delivery and storage of test materials to ensure they remain secure prior to the test session; arrangements at the test centre on the day of the test, including for example setting up the room, and carrying out pre-defined testing of computer hardware to be used in the computer-based components of the test; exactly how the test sessions are conducted, including for example procedures for checking on the identity of students turning up to sit the test, ensuring that each student is issued with the correct test booklet (the international contractor randomly assigns booklets to individuals on the lists of sampled students), what the test administrators are permitted to say to students who ask questions during a test session, and monitoring student behaviour during the test; the forms used to record attendance and to report other data from each test administration session; collecting, packing and shipping completed and unused test materials; and procedures for conducting any follow-up test sessions required to ensure response standards are met. Sometimes multiple visits to a school are required to reach the desired response rate of the designated sample of students.

As well as mandating these common and standard procedures, the international contractor also applies a system of quality assurance to monitor adherence to the procedures. Independently of the test administration system in each country, the international contractor employs and trains a small number of staff known as PISA Quality Monitors in each country, who attend a sample of test administration sessions to observe and record the procedures followed. The monitor prepares a report of each session observed, and highlights any discrepancies between the intended and implemented procedures. These reports are compiled by the international contractor and are used in the data adjudication process, a technical process undertaken by ACER and the PISA Technical Advisory Group to determine whether or not data from each country meet the PISA Technical Standards and are therefore fit for purpose.

Processing and Scoring Survey Responses

PISA survey instruments contain questions in a variety of formats. Some of them can be machine-scored. For example, responses to the various kinds of multiple-choice questions can be scanned directly into digital form, or they can be easily captured by data entry personnel and recorded digitally in the data processing system being used. No particular expertise is required to do this, and such processes are used to capture data from about a half the questions from the PISA 2012 cognitive instruments (the paper-based mathematics, reading, science and financial literacy questions in PISA) and from most of the questions used in the background questionnaires. PISA has implemented procedures designed to maximise the quality of data captured from these questions, including through the design of the response spaces and instructions given to student about how they should record their responses, and the provision of double data entry procedures as a quality assurance option taken up by several countries.

The major challenge presented at the stage of processing and scoring survey responses exists in relation to the (approximately) 50 % of items that require manual intervention to interpret the student response and convert it to a digital code. Ensuring quality and consistency in the way these responses are processed in the more than 60 countries that participated in PISA 2012 uses a number of steps.

During item development, possible responses to each question are identified, and these are categorised according to the level of knowledge of the variable measured in the question that is indicated by each response. Dichotomous items have two broad categories: those attracting *full credit* (for example the single correct answer to a multiple-choice question), and those for which *no credit* is warranted (the distracter response options). Some questions involve more than two response categories, so that particular responses may be of a quality that is clearly intermediate between the *full credit* and *no credit* categories. In these cases, a *partial credit* category can be defined. These ordered response categories, defined as part of the item creation process, are a critical part of each item. The response categories are described in the coding instructions for each item in relation to the particular knowledge and understanding needed for that category to apply, and the coding instructions also contain examples of particular responses given by students during item development, or in previous administrations of the item, to facilitate classifying observed responses into the defined response categories. The coding instructions for many of the released items have been published (see, for example, OECD 2009, 2013) and several are republished and discussed in Sułowska's Chap. 9 of this volume.

When the completed PISA test booklets are received for processing from each test administration centre (most commonly these are schools) within each country, after the field trial, and again after the main survey data collection is completed, teams of coders are assembled and trained to carry out the task of looking at student responses, assigning each response to one of the defined response categories for the item, and giving it the appropriate response code. Typically, teams of experts in each domain (for example, graduate students, or trainee teachers, or retired teachers) are recruited and trained by personnel from each PISA national centre to carry out this task. Those personnel had previously received training directly from the domain experts of the international contractor—indeed usually by the lead test developers in each domain—beginning a 'train the trainer' model. The contractor's domain leaders develop training materials that cover general issues in the coding of student responses, as well as specific issues in the coding of each item. They take the national coding team leaders through every item, teaching them how each kind of response should be treated.

Within each country, those team leaders pass on their learning to the local team they assemble. In Chap. 9 Sułowska, the leader of the coding team in Poland, describes her experiences in this role. Team leaders often develop additional sets of response examples to complement the material provided in the coding guides and through the international training. They implement local quality management procedures to ensure that all members of the national coding team are applying consistent standards as they work through the student material. In PISA 2012 an online coding process was introduced that was taken up in several participating countries, which used scanned images of student responses, and which allocated responses to members of the coding team in a systematic way. Typically the process involved coding all available responses to a particular item before moving on to the next item, in order to help focus concentration on the particular issues associated with each item, to rationalise coder training, and to remove the potential for bias associated with coder perception of the set of responses in a particular student's question booklet. Control scripts (student responses for which correct response codes were known in advance) were used periodically in the item allocation to monitor consistency of standards, to identify individuals who were not applying the standards correctly, and to identify items that were generating disagreement and therefore may have warranted additional training.

The international contractor provided an additional service to support national coding teams to complete their work. An international coder query service was implemented, whereby student responses found by coding teams to be difficult to classify, could be transmitted to ACER and the test developers could provide advice on the correct coding. Those responses were circulated among all national coding teams as a further means of achieving consistency of coding standards especially for hard-to-code items.

As a final check on the consistency of coding within each national coding operation and across the coding operations mounted in each country, ACER implemented formal coder consistency studies. At the national level, a random sample of student material was identified by ACER and national coding teams were required to have four coders independently code each selected item. The resulting data were analysed and reports were generated on the degree of consistency of output of each national coding operation. At the international level, a further sample of work from each national coding operation was identified by ACER for shipping to a central location, and an independent team coded the sample of work. Again, the data from this process were analysed to generate measures of the degree of consistency of output across participating countries. The studies to monitor coder consistency and the outcomes of these are reported in more detail in the various PISA technical reports (e.g. Adams and Wu 2003).

Data Capture, Processing and Analysis

The final stages of preparing PISA data for reporting lie in the steps of data capture, processing and analysis. Participating countries submit the data captured from the test sessions they conduct in purpose-built data capture software that the international contractor provides to all countries. The software ensures that the data entered into the various pre-defined fields meet the data definition requirements, and permit subsequent processing using a suite of analysis tools that are built for the purpose and applied across the entire dataset. The contractor's sampling experts first process the submitted data to ensure that the sampling plans were adhered to and that the data represent the population in accordance with the sampling variables defined earlier. A team of analysts at ACER check the data submitted by each country for each variable to ensure consistency and completeness, and engage in a dialogue with the data manager in each national centre to clarify any instances where the submitted data appear to lack consistency or are incomplete. A preliminary analysis of the data for each country is carried out, and detailed reports are generated and delivered to each country to provide an opportunity for analysts in each country to review their data and check any unexpected observations about the data. At that stage, data are sometimes identified that indicate an unacceptable degree of inconsistency, for example a particular item may have been unusually difficult in a particular country, or responses to particular questions in the background questionnaire may appear to be inconsistent, and possible explanations are then sought before a final decision about inclusion or exclusion of those data is made by the PISA Technical Advisory Group during the data adjudication process.

The analytic methods used in PISA are similar to those used in other large-scale surveys. They are designed to generate statistically the best possible estimates of the population parameters targeted by the survey. Those tools and techniques are described in a technical report published after each survey administration (e.g. Adams and Wu 2003). The OECD makes all of the resulting data publicly available for use by researchers and others.

Summary

The PISA survey is an enormous undertaking, involving the co-operation of a very large number of people in many countries. While the stakes are not high for individual students who participate, there is a growing interest in the kinds of comparisons made from PISA data, and the kinds of policy decisions that are taken by governments and education systems in response to PISA outcomes. As a result, PISA and its outcomes are increasingly exposed to public scrutiny. What comments and questions might we expect?

"PISA doesn't test the things we are really interested in, and the test questions they use do not match what students in our schools are taught."

"How can it be fair that PISA students in countries as disparate as Albania, Argentina, Australia, Austria and Azerbaijan undertake an assessment with test items written by test developers in Australia?"

"I heard that students in one particular country do well in PISA because in that country, only the very best students are chosen to undertake the PISA survey."

"In one country I know about, teachers help students with their PISA tests, students are allowed to stay in the test session until they have completed their test booklet no matter how long that takes, and then the teachers mark the students' examination papers very leniently."

This chapter was written to expose the myths such as those above that are voiced about PISA, and to answer legitimate questions that potential users of PISA data might have. The quality assurance mechanisms employed at each stage of the development and implementation of the PISA survey result in the generation of data that help to answer many questions about the state of educational outcomes across a large part of the globe.

References

Adams, R., & Wu, M. (Eds.). (2003). *PISA 2000 technical report*. Paris: OECD Publications.

Jago, C. (2009). *A history of NAEP assessment frameworks*. Washington, DC: National Assessment Governing Board. ERIC Document Reproduction Service No ED509382.

Organisation for Economic Co-operation and Development (OECD). (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2009). *PISA: Take the test*. Paris: OECD Publications. http://www.oecd.org/pisa/pisaproducts/Take%20the%20test%20e%20book.pdf Accessed 17 May 2014.

Organisation for Economic Co-operation and Development (OECD) (2013) PISA 2012 released mathematics items. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf Accessed 8 Oct 2013.

Rowe, H. A. (1985). *Problem solving and intelligence*. Hillsdale: Lawrence Erlbaum.

Chapter 7
The Challenges and Complexities of Writing Items to Test Mathematical Literacy

Dave Tout and Jim Spithill

Abstract The key to obtaining valid results from a large, international survey is having access to assessment items that are fit for the intended purpose. They must align with and incorporate the requirements of the relevant framework, give students fair and reasonable opportunity to demonstrate their true level of performance, cover a wide range of student abilities and mathematical literacy content, and work well in many different languages and cultural contexts. This chapter describes in detail the process that item writers from the PISA international contractors applied to generate items for the 2012 survey, from initial draft to final assessment, for both paper-based and computer-based items.

Introduction and Background

In PISA 2012 mathematical literacy was the major domain for the first time since 2003 so a comprehensive new set of items needed to be developed, including items for the new optional computer-based assessment of mathematics known as CBAM.

The mathematics development work for PISA 2012 was shared among seven different test development teams: the Leibniz-Institute for Science and Mathematics Education (IPN) and Universität Kassel both in Germany, Analyse des systèmes et des pratiques d'enseignement (aSPe) in Belgium, the Institutt for Laererutdanning og Skoleutvikling (ILS) in Norway, the National Institute for Educational Policy Research (NIER) in Japan, and the University of Melbourne and the Australian Council for Educational Research (ACER) both in Australia. Initial item drafts were also submitted by participating countries and these were all reviewed by the international test development teams and then the most promising of these developed for selection into the field trial and potentially the main survey. The lead international contractor for PISA 2012, ACER, oversaw the process and

D. Tout (✉) • J. Spithill
Australian Council for Educational Research, Camberwell, VIC, Australia
e-mail: David.Tout@acer.edu.au; James.Spithill@acer.edu.au

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_7

managed the item development teams and the review processes as well as the finalisation and preparation of the final survey instruments.

Because mathematical literacy was again the major domain, the item development process benefited from input from such a diverse consortium. In total, drawing from many more initial ideas, 345 new items were written for the paper-based assessment, of which 172 advanced to the field trial, and 72 of those were used in the main survey. The corresponding numbers for the computer-based assessment were 122 new items, with 86 used in the field trial and 41 of those used in the main survey.

The present authors, as test developers on this international assessment, quickly learned about the complex process and sophistication of developing and preparing suitable test items. It was a steep learning curve. It was not like a lot of test development where test developers sit at their own desks, write some good questions covering specified skills, submit them and then see them magically appear in a final assessment. An item developer in PISA soon learned that this was not the case.

This chapter focuses on the item development process, from the beginnings of an item, through revisions, to potentially ending up as an item in the main survey. The chapter attempts to describe the skills, knowledge and quality assurance processes that guarantee that the final survey assesses what it is supposed to assess. The test development process for PISA has to be particularly well developed because of the constraints of a large international assessment and the scrutiny to which the surveys are rightly subjected. For this reason, it was judged that this chapter should include general aspects of test development relevant to many assessments, as well as features specific to PISA.

The terminology that test developers use in PISA is that items begin with a real-world stimulus, which may be long or short (see, for example, the first sentence and image of Fig. 7.1). One or more questions then follow using the same stimulus material. The set of questions that derive from the same stimulus make up a unit. The unit PM942 Climbing Mount Fuji (OECD 2013b) shown in Fig. 7.1 has three questions. The word 'item' refers to the stimulus, the question, and the instructions for coding responses to the question.

Telling the Story: A Sample Unit

Throughout this chapter we will use one unit from the PISA 2012 survey to illustrate the process of item writing. PM942 Climbing Mount Fuji originated at ACER, and it has been chosen partly because it went through its full development in the hands of the authors and colleagues, ending up in the main survey of 2012. Figure 7.1 shows the final version of the unit. This chapter discusses the reasoning behind its evolution from the initial version shown in Fig. 7.2: how the stimulus text was tightened, how information in the stimulus was aligned more closely with its

CLIMBING MOUNT FUJI

Mount Fuji is a famous dormant volcano in Japan.

Question 1

Mount Fuji is only open to the public for climbing from 1 July to 27 August each year. About 200 000 people climb Mount Fuji during this time.
On average, about how many people climb Mount Fuji each day?

A 340
B 710
C 3400
D 7100
E 7400

Question 2

The Gotemba walking trail up Mount Fuji is about 9 kilometres (km) long.
Walkers need to return from the 18 km walk by 8 pm.
Toshi estimates that he can walk up the mountain at 1.5 kilometres per hour on average, and down at twice that speed. These speeds take into account meal breaks and rest times.
Using Toshi's estimated speeds, what is the latest time he can begin his walk so that he can return by 8 pm?

Question 3

Toshi wore a pedometer to count his steps on his walk along the Gotemba trail.
His pedometer showed that he walked 22 500 steps on the way up.
Estimate Toshi's average step length for his walk up the 9 km Gotemba trail.
Give your answer in centimetres (cm).

Fig. 7.1 Final version of PM942 Climbing Mount Fuji (OECD 2013b)

CLIMBING MOUNT FUJI

Mount Fuji is open for walking from the 1 July to the 27 August each year. About 200 000 people walk up Mount Fuji during this period each year.

Question 1

On average, about how many people walk up Mount Fuji each day during this period?

A 340
B 700
C 3400
D 7000

Question 2

Toshi took 7 hours to walk to the top of Mount Fuji along the Gotemba trail.
The trail is 9.1 kilometres long.
What was Toshi's average walking speed in kilometres per hour?
Give your answer to one decimal place.

Question 3

On his 9.1 km walk along the Gotemba trail, Toshi estimated that the length of each of his steps was about 40 centimetres.
Using Toshi's estimate, about how many steps did he take to walk to the top of Mount Fuji along the Gotemba trail?

Fig. 7.2 Initial version of the PISA 2012 unit PM942 Climbing Mount Fuji

relevant question, and how questions were significantly revised and restructured to better meet the intent and purposes of the PISA Mathematics Framework. The rationale for decisions about the CBAM unit CM013 Car cost calculator (ACER 2012) shown in Fig. 7.4 below will also be discussed.

PISA Test Development Process

An extensive process helps to guarantee the quality of the items. This is outlined in Fig. 7.3. ACER and the test development centres used a team approach to item writing, whereby experienced test developers wrote the items (with initial ideas from many sources) and met together to critique each other's items, following which the items were revised and improved. Then the revised items went through further comprehensive reviews and revisions. This included what ACER calls cognitive laboratories and pilots with potential test-takers, feedback from participating countries, and revisions made during a formal translation and review process with language experts. After the review processes, a field trial was undertaken with a sample of the target population in each participating country. The field trial data were analysed psychometrically and the results of this analysis guided the selection of the best performing items for the main survey. The final selection had to meet the criteria established in the Mathematics Framework (OECD 2013a), the technical requirements and the preferences as expressed by each country. The following sections elaborate on these strategies and processes. Some of these processes are also mentioned briefly in Turner's Chap. 6 in this volume.

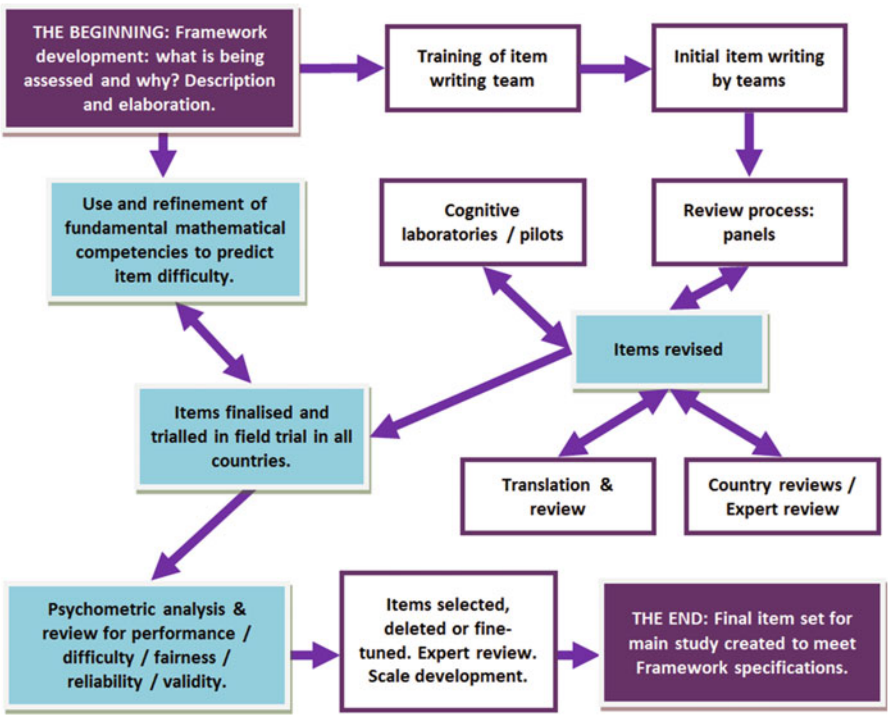


Fig. 7.3 ACER's test development process for PISA 2012

Before Item Writing Begins: The Conceptual Framework

Test development and writing proceeded from the agreed conceptual framework (OECD 2013a) that included a description of what was being assessed and why and how. The PISA Mathematics Framework for each PISA survey was developed by the Mathematics Expert Group (MEG), a team of international experts from different countries, drawing on scholarship described in Chap. 1 of this volume.

For test developers the Framework was the crucial document in that it established the requirements for the items to be developed. As noted elsewhere (see Chap. 1 of this volume), PISA is not a curriculum-based assessment. The PISA definition and description refers to the ability of the student to cope with tasks that are likely to appear in the real world, that contain mathematical or quantitative information, and that require the activation and application of mathematical or statistical knowledge and skills. This was the key challenge for test developers—to write items testing mathematical literacy and not just standard school-based mathematics.

The PISA Mathematics Framework (OECD 2013a) also specified the proportions of items with certain characteristics in the final survey (see also Chap. 1 of this volume). For example, approximately 25 % of items should be in the multiple-choice format and approximately 25 % should belong to each of the four mathematical content categories. Item developers needed to ensure that they provided the Mathematics Expert Group with sufficient items in each cell of the specification grid to allow for a good selection of items for the main survey.

The Item Writing Process

Item writing for PISA proceeded through the stages outlined in the diagram in Fig. 7.3, and depended on a wide range of knowledge, experience and skills. This section outlines the formal processes and mechanics of item writing that were followed, and also its more creative and challenging aspects.

Test Development Teams' Induction and Training

Before writing commenced for PISA 2012, key members of each of the test development teams met and were introduced to the PISA Mathematics Framework and trained in the item writing process, including the mechanics and quality assurance processes, and approaches to writing successful items. Item writers in

PISA needed to meet different cultural and linguistic demands, to address the various requirements and specifications in the PISA Mathematics Framework, and also to address specific new requirements and expectations for PISA 2012. A similar training session was also provided to the National Program Managers and relevant country personnel to support countries that intended to submit potential items.

Based on feedback and reactions to previous PISA assessments, some specific key challenges were set for PISA 2012 mathematical literacy test developers. These included that the suite of new items should:

- be more realistic and authentic than items in previous surveys, which had been produced for PISA 2003 when mathematics was first the major domain and test developers were themselves coming to terms with the relatively new notion of mathematical literacy
- make the contribution of school mathematics content more explicit and more easily recognisable to external observers than in some items of previous surveys
- include a greater number of more difficult items that allow capable students to demonstrate their ability
- include a greater number of very easy items so that the level of performance of students at the lowest levels could be better measured.

Examples of the impact of these requirements can be seen in the revisions made to the unit PM942 Climbing Mount Fuji. The changes made to Questions 2 and 3 were explicitly made to make the questions more authentic. Also in the computer-based unit, CM013 Car cost calculator (ACER 2012) shown in Fig. 7.4 below, there were a number of questions developed to meet the second and third requirements in the above list—to make the contribution of school mathematics content more explicit and more easily recognisable, and to include a greater number of more difficult items.

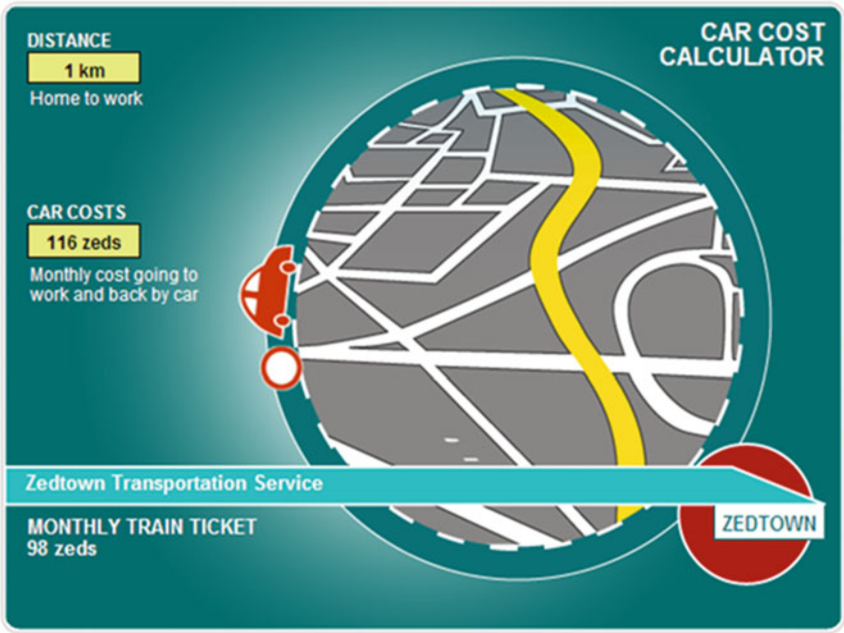
Optional Computer-Based Assessment

In PISA 2012, an optional computer-based assessment of mathematical literacy (CBAM) was introduced for the first time. In CBAM, specially designed PISA units are presented on a computer, and students respond on the computer. They are also able to use pencil and paper to assist their thinking processes. The CBAM initiative is further discussed in Chaps. 1 and 8 of this volume.

This required a new set of skills for the test development teams, as the CBAM option provided opportunities for test developers to write items that were more interactive and engaging, and which may move mathematics assessment away from the current strong reliance on written, text-based stimuli and responses, potentially

Car Cost Calculator CM013Q03 (format changed from original)

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.



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.

Fig. 7.4 CBAM item CM013Q03 Car cost calculator Question 3 (ACER 2012)

enabling different student abilities to be assessed. The challenge posed to both the test developers and the computer platform development team and programmers was to make CBAM more than a version of the paper-based assessment transferred onto a computer. The intention was to develop items that reflected the real-world use and application of mathematics within a computer-based environment, but also to take advantage of the potential to assess aspects of mathematical literacy that could not be assessed with paper-based assessment. The styles and types of items and interactivity included: drag-and-drop items; the use of hot spots on an image to allow students to respond non-verbally; the use of animations and representations of three-dimensional objects that can be manipulated; the ability to present students with sortable datasets; and the use of colour and graphics to make the assessment more engaging.

With the above in mind, a classification scheme was developed by the ACER test development team to classify the items that were developed. The non-mutually-exclusive categories described were:

- animation and/or manipulation
- automatic calculation, where calculation was automated ‘behind the scenes’ to support assessment of deeper mathematical skills and understanding
- drawing, spatial, visual cues and/or responses
- automatic function graphing and statistical graphing
- simulation of common computer applications (e.g. using the data sorting capability of an ‘imitation’ spreadsheet)
- simulation of web-based applications or contexts, with or without computer-based interactivity (e.g. buying goods on line).

The following sections in this chapter apply to the writing of both paper-based and computer-based mathematical literacy items. However, developing CBAM items posed additional challenges especially as this was the first time this assessment was offered. At first, the computer platform was still under development, so it was unclear what interactivities would be supported and not all the envisioned interactivities were eventually realised. For example, there was no ability for students to enter mathematical symbols (apart from the standard set of key-board symbols), expressions or formulae into the system. The use of video or audio was not practical, especially because of the large number of languages. Many of these limitations arose from the complexity of providing a platform that could be used in a large number of countries around the world, using equipment that a random sample of schools were highly likely to possess at that time, and supervised by test administrators without special computer expertise. The screen size (the available ‘real estate’) restricted the number of words and images. It was necessary to allocate extra space in the English source versions of each item to allow for the longer text forms that occur in many other languages. The design process used mock-ups and story boards of the items, and interactive items were sometimes programmed initially in Excel or Word so that a meaningful item review process could be undertaken cost effectively. Item writers had to work hard to communicate their vision to the programmers, illustrators and designers.

The Challenges and Complexities of Item Writing: A Creative Science or Art?

When writing PISA mathematical literacy items, there was no single fixed process to follow. There were certainly a number of processes and structures available that supported the item writing process, and these are explained in more detail in this section. The challenge for item writers was to create items that:

- were rich and interesting for 15-year-olds around the world and were neither too hard nor too easy
- had obvious authenticity and did not pose seemingly artificial questions
- were as much as possible equally accessible and equitable for students of different gender, culture, religion, living conditions
- used appropriate and accessible language.

Where and How to Begin?

One of the key creative aspects was to find a context with realistic and authentic mathematical content likely to be accessible to and engage 15-year-olds across the world. One approach was to start with a real-world context and develop it into a unit. The problem with this was that often the real-world context was too complicated and complex for 15-year-olds in a test situation. Often the mathematics was too highly embedded in the context and to extract the mathematical model required too much reading and understanding of the situation, which would block many students from solving the problem. Another issue was that the mathematical formulas and the required quantities or numerical information to be manipulated in the real-world context were also complex and so calculation would be time consuming and open to arithmetical errors, thereby clouding what the item assessed. It was important to simplify the real-world context, the related stimulus and its embedded mathematical information to make it accessible whilst still maintaining the authentic aspect.

The CM013 Car cost calculator unit (ACER 2012), the stimulus and one item of which is shown in Fig. 7.4, is an example of a unit that began from a real world experience. The idea was stimulated by a cardboard calculator handed out freely by a transport authority. It then developed into a CBAM unit because the real manipulative cardboard calculator stimulus could not be used in an international paper-based assessment. The electronic version also had strong face validity: many websites have similar features. This is an example of how online assessment extends the range of authentic situations that can be used.

The interactive car cost calculator could be manipulated by the student to see what impact the distance variable had on the cost of car travel and to gather data for answering a number of questions. This allowed the student to focus on the functional relationship between the variables rather than using a formula or table of values. This unit would have been much more difficult to write as a realistic paper-based item, given the inability to 'hide' the formula behind the scenes.

Another approach to item development was to start with a mathematical concept or content area and try to find an appropriate context based on an authentic real-world task. The problem with this approach was that often this resulted in what is traditionally seen as a school, curriculum-based, word problem that has little real-world relevance or authenticity. Many of the items submitted by countries were of this style, and few such items were able to be developed for use in the PISA main survey.

An idea for a unit often developed from a test developer's personal experiences or interests, or from something they read or found—in the media, in the outside world, at home, in the community or in a workplace. In other cases, often still based on such an observation or interest, a test developer searched on the internet for related examples or contexts that would be a suitable starting point and then turned that into a useful context for asking mathematical literacy questions suitable and relevant to 15-year-olds. The unit PM942 Climbing Mount Fuji is a good example. The test developer was looking for a context that he could use to develop a unit about walking (his personal hobby) to assess skills related to speed, distance and time relationships. Units are likely to be more authentic and accurate if test developers write about the things they know. Because he was aiming to engage an international audience, he chose Mount Fuji as an iconic physical feature that many students would know. Although not having personally walked Mount Fuji, the writer was able to select and evaluate the information he found when researching—he knew what he was looking for. In this sense the item writer could guarantee that the context and the related mathematics were realistic and dealt with factors that really have to be considered by walkers.

Some items started small and grew, while others started as a big idea that was edited and reduced to suit the 15-year-old test taker. As mentioned above, simplification of the context and related stimulus was usually needed. It can often be useful if the item writer has in mind a particular student they have taught when trying to set an item at an easier or harder level within the overall set of items. Sometimes to fill gaps in the item set, a test developer was required to write a unit fitting Framework specifications e.g. to write items for a specific content category (e.g. *Change and relationships*), context category (e.g. *Occupational*) and process (e.g. *Employ*, perhaps using a formula) from the Framework (OECD 2013a).

Use of Visual Support

No matter which approach was used to develop the unit, there was always the need for some form of visual support for the stimulus. This had been the case with earlier PISA surveys, but was seen as a feature to be strengthened for the PISA 2012 survey, where there was a more extensive and consistent use of visual support by the use of illustrations, diagrams, or photographs. This was used to increase accessibility of the problem, by tuning the student in to the context and thereby helping to reduce the reading demand. In other words, the visual support helps make the unit attractive to students and helps connect the content to the real world and give the questions a purpose.

The Mechanics of Item Writing

Item writers were expected to provide a variety of items that met the framework specifications for context, format, content, processes and fundamental mathematical capabilities as described in Chap. 1 in this volume. The list below gives a number of requirements that needed to be operationalised through the test development process.

- There needed to be a full range of difficulty so that all participating students would find some items that gave them an opportunity to demonstrate what they could do.
- A requirement of the psychometric model is that items should be independent of each other to the maximum extent possible. In particular, a response to one item in a unit must not be required in solving another item.
- Items should not require excessive computation. Whilst items could include computations (as they might naturally arise in the context), the items were generally not to test great computational dexterity.
- The level of reading required should not interfere significantly with a student's ability to engage with and solve an item. Practical guidelines were issued for this.
- No single item should take more than five minutes to complete, and no unit more than 15 min so that students had sufficient time to attempt a range of independent items. This is needed for the psychometric model. This criterion led to a number of interesting items being discarded before the field trial.
- Items were to be culturally acceptable across participating countries, and should be readily translatable.
- Student responses must be able to be consistently scored (coded) in an efficient manner by teams around the world.

A standard Word template was provided so that all item writers wrote to the same style and format. The template ensured that the item metadata was a consistent reflection of the Framework. The template included a section for the coding of each item (for further details see Sułowska's Chap. 9 of this volume), a question intent description and the Framework process, content and context categories. Figure 7.5 shows the basic coding instructions of PM942Q02 Climbing Mount Fuji Question 2. This information is provided for all newly released items (e.g. OECD 2013b). The question intent is a brief description of what the student has to do to solve the problem. For coding, the item writer needed to specify the types of response that would receive Full Credit, Partial Credit (where relevant) or No Credit. The template ensured that all these issues were addressed by the item writer and discussed, reviewed and agreed upon in panel sessions.

CLIMBING MOUNT FUJI SCORING 2		
Question Intent		
	Description	Calculate the start time for a trip given two different speeds, a total distance to travel and a finish time
	Content	Change and relationships
	Context	Societal
	Process	Formulate
Full Credit		
	Code 1	11 (am) [with or without am, or an equivalent way of writing time, for example, 11:00]
No Credit		
	Code 0	Other responses.
	Code 9	Missing.

Fig. 7.5 Scoring and Question Intent section of PM942Q02 Mount Fuji Question 2

The coding scheme for this question shown in Fig. 7.5 recognises that a student who has arrived at a correct value of 11 has achieved the question intent and is not penalised for omitting the time specification 'am' from their response. This is a point of difference between mathematical literacy and common school mathematics teaching practice, where teachers may well deduct marks if such information is not written along with the numerical answer. In PISA it is a case of giving credit for what a student can do. Teachers aim to develop good habits in their students, which is different to the measurement aims of the PISA assessment.

The Metadata

Test developers must map each unit and item against the characteristics of the PISA Mathematics Framework. These item characteristics become metadata for each item. For PM942 Climbing Mount Fuji, the key item characteristics (metadata) for the items, for both the final version (Fig. 7.1) and the initial version (Fig. 7.2) are shown in Table 7.1. The process categorisation only occurs in the final version, because initial test development began before this new aspect of the Framework had been finalised. The estimated difficulty was obtained by test developers by rating against the fundamental mathematical capabilities, as described below.

PM942Q02 was completely redesigned between the initial and final versions and this increased its estimated difficulty substantially from 4 to 10. In the initial version (see Fig. 7.2), time and distance were given directly, so the student had the straightforward task of making a single calculation to find the average speed. In contrast, the final version (see Fig. 7.1) demands two different time calculations based on related speeds and then a calculation of a latest starting time, where even the notion of 'latest start' is linguistically not simple for many students.

Table 7.1 Item characteristics of final and initial versions of PM942 climbing Mount Fuji

Item characteristics	Question 1	Question 2	Question 3
Final version (see Fig. 7.1)			
Mathematical content	Quantity	C&R	Quantity
Context	Societal	Societal	Societal
Process	Formulate	Formulate	Employ
Estimated difficulty	5	10	9
Response type	Multiple choice (simple)	Constructed response expert	Constructed response manual
Initial version (see Fig. 7.2)			
Mathematical content	C&R or quantity	C&R	C&R or quantity
Context	Societal	Societal	Societal
Process	Processes category definitions not finalised until after this stage		
Estimated difficulty	4	2	4

C&R Change and relationships

Item Format and Item Response Types

PISA had used items with different presentation formats and with a range of response format types for the earlier paper-based surveys, and these item format types were combined with new presentation formats developed for PISA computer-based assessments in 2009 (science and electronic reading) and 2012. The item response categories described and used for PISA 2012 were:

- Constructed Response Expert—items where the student writes a response that needs expert judgement for the coding. In PM942Q02 Climbing Mount Fuji Question 2, the field trial data indicated students could sometimes add in comments and valid variations. The complex coding process for these items is described by Sułowska in Chap. 9 of this volume. Expert coded items are often intended to measure higher level thinking, argument, evaluation and the application of knowledge, and they might involve constructing mathematical expressions or drawings and diagrams that necessitate the involvement of a suitably expert person to assign observed student responses to the defined response categories.
- Constructed Response Manual—items that have a very limited range of possible full credit responses (e.g. single number or name) but are best coded manually

although extensive training is not required. PM942Q03 Climbing Mount Fuji Question 3 (single number response) is an example. Items of this type work well in place of a multiple-choice format that has too many or too few good distracters, and they reduce the potential for guessing.

- Constructed Response Auto-coded—items that can be automatically coded. The actual response is keyed in by a data entry operator as part of the processing of responses, or in the case of computer-based items captured directly by the computer. Many CBAM items were of this type, including CM013Q03 Car cost calculator Question 3 (see Fig. 7.4).
- Simple Multiple Choice—items where there is one correct response that the student selects (e.g. PM042Q01 Climbing Mount Fuji Question 1). This includes both radio buttons and a drop down menu where there is a unique correct auto-coded response.
- Complex Multiple Choice—items where the student responds to a set of multiple choice statements (usually two or three) and selects one of the optional responses to each (for example, 'true' or 'false'). The item is only coded correct if all responses are correct. Items of this type could be automatically coded and were used in both the paper-based and computer-based assessments. They reduce the effect of guessing.
- Selected Response Variations—these variations to the standard multiple-choice formats above were only used in CBAM, and could all be coded automatically.

Preparing for Reliable Coding

In constructed response items the challenge for the item writer is that the question stems need to be well structured with clear instructions to the student, as in PM942Q02: 'Using Toshi's estimated speeds, what is the latest time he can begin his walk so that he can return by 8 pm?' Even with clear instructions, there are many ways in which the student could write the time (e.g., 11 am, 11:00, 11 in the morning, 11) and so manual coding of the responses is required. Because of this, the item writer also needs to communicate explicitly with the coder through the coding guide. The potential range of responses needs to be anticipated and then documented fully for reliability and ease of coding. Further examples are discussed by Sułowska in Chap. 9 of this volume.

Use of the Fundamental Mathematical Capabilities

Test developers estimated the item difficulty of each item before the empirical data of the field trial was available using their professional judgement based on their experience of students generally and in the cognitive laboratories in particular, and also created a score (see Table 7.1) by rating the items against the fundamental mathematical capabilities, as described by Turner, Blum and Niss in Chap. 4 of this volume. This procedure predicted for the Climbing Mount Fuji unit, that Question1 (PM942Q01) would be much easier than Question 2 (PM942Q02) or Question 3 (PM942Q03). In the field trial, the success rates across all countries were 46 % for PM942Q01, 12 % for PM942Q02, and 11 % for PM942Q03 (full credit) and a further 4 % with partial credit. This shows that using the rating scheme did predict difficulty quite well and also that quite difficult items had total scores much below the theoretical maximum rating of 18. The other use of the fundamental mathematical capabilities was to ensure that the sets of selected items were balanced across different aspects of mathematical literacy. Additionally, questions could be devised to highlight *Reasoning and argument* or *Using symbolic*, *formal and technical language and operations* over other capabilities to round out the item set.

The Three Processes

An issue that affected test development was the determination, after some item writing had commenced, to apply the new classification of items against the three processes of mathematical literacy developed in the 2012 revision of the Framework as explained by Stacey and Turner in Chap. 1 of this volume:

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

For example, in PM942Q01 Climbing Mount Fuji Question 1 (see Fig. 7.1) the main cognitive demand on students was to understand the problem and its real-world meaning in order to recognise that they could use the dates to work out the number of days that Mount Fuji is open, and divide the total number of people by this number. This meant that it fell predominantly into the *Formulate* process. For 15 year old students, there was lower demand from the *Employ* (the calculation) and *Interpret* processes. In contrast, in PM942Q03 Climbing Mount Fuji Question 3, the mathematical process required was much more explicit and matched a standard process for conversion within metric units. This item was hence classified as *Employ*. As the test developers and the Mathematics Expert Group applied the new classification to items from earlier PISA surveys, it emerged that it was not always easy to draw sharp lines between the processes and the classification hinged on what was judged to be the main cognitive challenge or impediment to a student solving the problem. When writing new items for PISA 2012 after the *Formulate—Employ—Interpret* classification definitions had become available, test developers constructed items that focused more strongly on just one process. Within the Mathematics Expert Group, there was agreement that tasks that best encapsulated mathematical literacy in its fullest sense would usually involve aspects of all three processes, since they reflected all stages of the mathematical modelling cycle.

Classroom activities that use tasks like this are required to develop mathematical literacy to the full. However, in the context of an assessment, items that measure abilities in constituent processes are valuable. In Chap. 11 of this volume, Ikeda discusses this issue more fully.

Review Processes

The Panel

After individual test developers drafted items they met as a panel (at least three writers) to critique and review each other's items. Panel members individually examined the items before the panel meeting. Questions addressed during the panel included the following:

- Is the mathematics correct?
- Does the content sit well with the PISA Framework?
- Does the metadata accurately describe the item against the Framework criteria?
- Is each question coherent, unambiguous and clear?
- Is it clear what constitutes an answer: do students know exactly what they should produce?
- Is each question self-contained? If it assumes prior knowledge, is this appropriate?
- Are there dependencies between items (e.g. does one item give a clue for the next one)? Would a different order of items with the unit help or hinder students?
- Is the reading load as low as possible? Is the language simple and direct?
- Are there any 'tricks' in the question that should be removed?
- Are the distracters for the multiple-choice items plausible, or can better distracters be devised?
- Are the response categories complete and well defined, and is the proposed coding easy to apply?
- Is the context appropriate and relevant for the target group?
- Is the context authentic?
- Are the text and the questions fair? Are there any ethical matters or other sensitivities that may be breached (for example, racial, ethnic, gender stereotypes, and cultural inappropriateness)?
- Do the questions relate to the essence of the stimulus?
- Is the proposed scoring consistent with the underlying ability that is being measured? Would students possessing more of the underlying ability score better on this item than students with less?
- Is it clear how the coding would be applied to all possible responses?
- Could partial credit be given if part of the answer is achieved?
- Are there any likely translation difficulties?
- How would this item stand up to public scrutiny?

Two to three hours were allocated for each panel to discuss 20 items. Discussion within a panel meeting was direct and robust, and each member was expected to comment on each item. Virtually no items escaped amendment of some kind. How this process impacted on Climbing Mount Fuji is discussed in a later section. Immediately after a review, resulting changes were implemented in the PISA item development database while the discussions and amendments were fresh in the test developer's mind. The new versions were then cross-checked by another panel member to ensure that they met the panel recommendations.

Some items required only minor changes, such as splitting long sentences with conditional sub-clauses into shorter, more direct statements, using active rather than passive voice, or moving stimulus material from the beginning of the unit to be adjacent to the relevant item, and editing diagrams. Other items required major changes. Items that could not be reworked to the satisfaction of the panel were discarded. Such items tended to be not realistic, too difficult for the target audience or too time consuming to solve. Sometimes a panel suggested an additional item to complement a unit. In some multiple-choice items the distracters were judged weak or artificial but the mathematics itself was interesting and sound. Such items could be changed to a constructed response item or multiple-choice options could be improved. In the case of complex multiple-choice items, sometimes the item writer supplied three or four multiple-choice statements all of which went to the field trial with a decision afterwards to retain all or delete one or more. By using trial data on each statement, the difficulty level and the other psychometric properties of the complete item could be manipulated. This is one of the few item-level modifications that could be safely made after the field trial. Everything in the main survey needs to have been tested in advance.

Student Feedback from Cognitive Laboratories and Pilot Study

The items were also tested with local students of the target age group in cognitive laboratories in the early stages and in later larger pilot work. A cognitive laboratory involves a test developer meeting with three or four students, observing how they work with the items and then discussing with them any issues that affected their interpretations or approaches. Essentially this uses the long established 'think aloud' interview methodology commonly used in mathematics education research (Ginsburg et al. 1983). Student responses were then used for reworking or discarding items.

The test developer explained first that it was the items that are being tested, not the students. Students were given one unit at a time with the test developer observing how they went about working it out and asking questions about their actions and reasons. When all members of the group were finished there was group discussion and feedback. Each student was asked to respond to these issues:

- whether it was easy to follow the instructions
- whether the content or context was familiar
- whether the content was difficult or easy
- whether the unit was interesting or boring
- specific comments about the stem, the distracters in multiple-choice items, the language used, and diagrams
- (for CBAM) ease of interactivity and navigation.

In a one hour session it was possible to cover about four or five units. For CBAM items, feedback often resulted in improved and simplified instructions and improved graphic design for the interactivity and navigation around the screen. The quality of the feedback varied, of course, but there were numerous cases where students were commended and humorously advised that a career as a test developer might well await them, given their insight into the testing process.

The pilot study involved more than 1,000 students in 46 schools across Australia where the lead international contractor (ACER) is located. These schools were not reused in the Australian sample for the field trial or main survey. Students worked through the near-final versions of the units allocated to 19 test booklets. The responses were analysed to check that the items were behaving as expected. Constructed responses were checked for the range of responses, expected and unexpected. Experienced coders from other ACER teams also coded the responses, and made valuable comments to simplify and clarify the coding guides. The students' responses were also used as examples within the coding guide and for the coder training workshops.

Country Reviews

National Project Managers from OECD member countries and partner countries and the Mathematics Expert Group reviewed items batch by batch and all their feedback was considered by the development team. For each item, reviewers rated each of the following criteria on a five point scale:

- What is the item's relevance to preparedness for life?
- How well does the item sit within the curriculum expectations for 15-year-old students in your country? (Although PISA is not curriculum based, it is necessary that items can be solved using mathematics that students have learned).
- How interesting is the item?
- How authentic is the context?
- Are there any cultural concerns with the item?
- Do you foresee any translation problems with the item?
- Are there any coding concerns with the item?
- Does the stated question intent reflect the content of the item?

Table 7.2 National program manager ratings for PM942 climbing Mount Fuji

Mean scores[a]	PM942Q01	PM942Q02	PM942Q03
Relevance to preparedness for life	4.32	4.49	4.23
Within the curriculum	4.57	4.49	4.49
Interest level	3.83	3.94	3.83
Authentic context	4.36	4.45	4.11

[a]Range 1–5, with 5 best

Table 7.2 summarises the feedback from 47 member and partner countries for the PM942 Climbing Mount Fuji unit (near final version) for the first four criteria above. The mean scores showed that Climbing Mount Fuji was highly regarded as a strong candidate for inclusion in the field trial. Fewer than 10 % of countries reported concerns with the unit under these criteria. For PM942Q01 the concerns were about the format of the date, and the interpretation of the phrase 'On average, about . . .'. For PM942Q02 there were comments about this being more of a science curriculum topic than mathematics in their country, and comments about authenticity noting that walking distances are often expressed in hours, not kilometres. For PM942Q03 the concerns included whether it is necessary to insist on stating the answer in centimetres, and that pedometers are not common so their function should be explained. This feedback was used in the revision of the items, and in the translation notes to allow customisation to local conventions, such as for representing time.

Translation Issues

The PISA 2012 survey was conducted in 39 different languages, so the need for translation into those languages impacted on the wording and structure of units and items. Through the comprehensive translation process and review system described briefly by Turner in Chap. 6 of this volume, language structure, the meaning of items, content and cultural issues are identified and addressed as an important thread within the item development process.

The translation process required dialogue between developers and the Linguistic Quality Control Agency (cApStAn) in Belgium under the guidance of the translation expert, also based in Belgium, engaged by ACER to oversee this process and to provide definitive advice on technical matters related to the preparation of national versions of each item. Agreed French and English 'master' versions of each item were constructed for translation into local languages. The French translation manager described how "English is concise but French is precise". The rewording in English that was often required to facilitate an unambiguous translation into French often helped make the English clearer.

Over the many months of this overnight email dialogue between Belgium and Melbourne a set of standards on structure and wording of items emerged. Some examples were avoiding truncated stems for multiple-choice items, stating units in each option rather than in the question stem, and accommodating how some languages (e.g. Slavonic) treat plurals. Item writers needed to accommodate vocabulary differences (e.g. some languages do not have a word for ‘million’, some languages use different words for concepts depending on the context, such as the area of a shape versus the area of a country where the same word ‘area’ is used in English). With CBAM units there was a limited amount of screen space, and sentences that fitted that space in English may not fit after translation: writers had to allow about 50 % more space for the translation than was needed for the original English version. Graphics and other layouts had to accommodate languages that are written right-to-left. Turner in Chap. 6 of this volume gives other examples.

The translation notes that accompanied each item specified what could be changed. According to local usage, translators routinely changed the decimal point or comma, used local conventions for operator symbols such as ÷ or / for division, and for writing dates and times. Translation notes specified when it was appropriate to change the letters used for algebraic variables in formulas to agree with the initial letters of the corresponding words (e.g. in $F = ma$ the letters are the initial letters of force, mass and acceleration) and whether to change metric to locally used units. These translation notes are included in the released versions of items (e.g. OECD 2013b).

The Impact of the Review: Climbing Mount Fuji

Comparing the final (Fig. 7.1) and initial (Fig. 7.2) versions of PM942 Climbing Mount Fuji shows a number of changes. The initial stimulus contained some key information that was moved to Question 1 where the data were needed. The data in the stimulus were replaced by a short scene-setting sentence. In the final version, the information for each item is presented within that item, which reduces the reading demand. The graphic design team produced an illustration to make the context more explicit and engaging, and experiences in the cognitive laboratories indicated that illustrations did indeed have this effect. The words ‘walking’ and ‘walk’ were changed at the panel stage to ‘climbing’ and ‘climb’ in Question 1, to be consistent with the unit title. Streamlining language in this way, so that different words are not used for the same idea, improved student comprehension and simplified translation.

Question 1 initially asked for an approximate answer rounded to the nearest ten, hundred or thousand. The panel standardised the options at two significant figures. The original option D 7,000 could have been the result of rounding either 7,143 (200,000 ÷ 28, the sum of the dates given 1 + 27) or 7,407 (200,000 ÷ 27, from ignoring the fact that the time period is over 2 months). This observation led to the addition of distracter E in the final version. To simplify reading and translation, the phrase ‘during this period’ was deleted because there is no other period during which the trail is open to walkers.

The initial version of Question 2 was thought to be too much a straightforward school exercise where the context was not really needed, and hence it was not in the PISA mathematical literacy style. There were also concerns from students in cognitive laboratories about the item not addressing real-world considerations such as meal breaks and rest times, and these considerations, when included in the thinking of students analysing the problem at a more sophisticated level, caused wrong answers. So the item was reworked into a more realistic scenario about planning a walk up and down the Gotemba track, with different average speeds when walking up or down the mountain, and then requiring the student to find the latest time to begin the walk. In order that the mathematical reasoning rather than the numerical calculations would provide the main challenge of this question, and to give a whole number response that would be easy to code, the distance was realistically rounded to 9 km, and the walking speeds were given as values that lead to whole number time calculations: $9 \div 1.5 = 6$ and $9 \div 3 = 3$. It was also made explicit that the total distance travelled was 18 km, so students would not be confused by the one-way trip of 9 km in Question 3. The careful, but still realistic, choice of values enabled the intent of the question to be met, with the focus on the *Formulate* process.

The original Question 3 of Fig. 7.2 met similar criticism of being artificial. In the real world a walker who was interested in their number of steps would most likely have a pedometer to count their steps. So Question 3 was turned around to give the two most easily known pieces of data, total distance and total number of steps, then asked for an estimate of average step length. The initial version had ‘to walk to the top of Mount Fuji’ in Questions 2 and 3 but ‘walk up Mount Fuji’ in the stimulus. It was decided to take out the reference to ‘top’ and refer to ‘up’ and ‘down’ throughout the unit. This streamlined the instructions, but also attended to a student concern expressed in cognitive laboratories about people who might take the walk but not make it all the way to the top.

In summary, the changes to Climbing Mount Fuji aimed to:

- make the unit realistic so that students could relate to the story, thereby helping them to link the different questions as they worked through them
- be consistent and direct in the use of language, and be specific about measurement units required for constructed response items
- remove unintended complications and ambiguities (e.g. whether meal breaks needed to be added) by addressing them explicitly in the text
- make the calculations straightforward so that the unit and items could focus on the mathematical literacy skills and processes being assessed.

From Field Trial to Main Survey

The field trial was the key winnowing stage for items. The MEG selected just over 180 new paper-based items and 90 computer-based items for the field trial, using a variety of information sources including detailed feedback from National Program

Managers, feedback from the ACER pilot study, independent reviews by each MEG member and the Mathematics Framework specifications.

After the items had been through their final proofreading, design and desktop publishing processes, they were grouped into clusters for trialling with a sample of the target population in every participating country. Each booklet in the paper-based assessment or each online CBAM test form was made up of a number of item clusters. These clusters were then rotated among booklets and forms. In the paper-based assessment each cluster consisted of about 12 items and in CBAM each cluster had about 10 items. After creating the clusters, the final mathematics field trial conducted in 2011 used 172 new paper-based items and 86 computer-based items.

Psychometric Review of Item Performance: Difficulty, Fairness, Reliability, Validity

The psychometric data from the field trial that summarised the measurement properties of each item were crucial for selecting items for the main survey. By the time the statistical review had to be finalised for the 2012 survey, results from approximately 6,200 students from OECD countries and additional students from partner countries and economies were available for each item. Many more students were involved overall, because individual students only complete a small number of the items. The set of items selected for the main survey had to satisfy several requirements, such as showing a good spread of difficulty. Every item was checked to see if it performed at the test developer's expected difficulty level—a significant deviation from what was expected could indicate an issue with the item, such as unexpected ambiguity. Test developers checked that the coding and scoring worked well. Each distracter for a multiple-choice question needed to attract an appropriate number of test takers.

Rasch scaling was used to calculate the item difficulty, and these item difficulties were used to obtain the required range of difficulty of the main survey items. Various statistics tested the fit of the item to the Rasch model. The correlation of the item score with students' scores on all the other items combined was calculated to indicate whether the item measured the same underlying ability as the survey as a whole and also contributed something unique. The ability (according to the Rasch model) of the students respectively with correct and incorrect answers for each item was calculated, to test that the average ability of students answering each item correctly was higher than the average ability of students providing an incorrect response. For example, in PM942Q02 Climbing Mount Fuji Question 2 successful students had an average ability of 0.39 (above the mean of 0) and unsuccessful students had an average ability of -0.83 (below the mean). Students who omitted the item had an average ability of -1.19. Statistics for each multiple-choice option were also analysed. Together, these item statistics indicated that this item and its multiple-choice options were working validly and reliably.

Sometimes students who know more or think more deeply do not score as well on some items as students taking a naïve approach. For each response code, the 'characteristic curve' of the probability of success of the item against student ability (as estimated from the Rasch model) was plotted and compared to the theoretical curve for each item. This gave guidance on how the item performs across the ability range, and picked up instances where more capable students read more into the problem than was expected by the item writer, or where there were unforeseen ambiguities or likely misinterpretations. As noted above, a potential instance of this was identified in the cognitive laboratories for PM942Q02 (then eliminated), when students who thought more deeply about the context allowed for meal breaks. The reworded item, modified to eliminate this potential ambiguity, performed well at the field trial.

Statistics also allowed examination of the performance in and between individual countries, in order to identify items with a cultural or linguistic bias or major mismatch with local curricula. There were no countries where PM942Q02 was significantly easier or harder than expected on the basis of the total scores. The item had low discrimination in only one country and higher than expected in only two countries and, in all countries, successful students had higher ability as measured by the whole item set than unsuccessful students. The reliability of coding was also examined as was the gender difference. Large gender difference may indicate cultural bias. Items that did not perform well on any of these psychometric measures were rejected, as there was no opportunity to adequately trial an amended item.

The Main Survey Items

The final selection of items for the main survey was made using all the data from the field trial at the MEG meeting in Melbourne in September 2011, also attended by ACER project managers, lead test developers, psychometricians and a representative of the secretariat of the OECD who ensured that the Framework criteria were implemented. For the paper-based assessment a large number of suitable items survived psychometric scrutiny from the field trial and were available for selection for the main study. For the CBAM assessment, which was designed to be smaller, a more restricted set of items was available because a much smaller set had been developed and trialled. As well, OECD had employed a separate organisation, *Achieve* (www.achieve.org), to conduct an independent validation and review process. At this MEG meeting, the *Achieve* external reviews of each item were made available to the MEG. Officers of *Achieve* reported that their reviewers had found the items in the new pool to generally be an improvement over previous surveys. The report cited one reviewer who noted that

> the present selection of items and the formulation of the questions are much better than in previous years, where the questions often were loaded with unnecessary—and [hard to read]—information. (Forgione and Saxby 2011, p. 17)

For the paper-based item set, the *Achieve* review was generally very positive. For the CBAM items *Achieve* reviews commented in a number of cases that there was little or no significant mathematics in the items. This led to discussion about the role and purpose of the CBAM items. By design, mathematical operations including calculation were often automated in CBAM (e.g. by the CM013 Car cost online calculator), so that the assessment could focus more on the *Formulate* or *Interpret* processes of a problem without being confused with calculation or substitution into a formula for example. In the view of the test developers this was a strength of CBAM but this view was not shared by all the *Achieve* reviewers. Some comments by the *Achieve* reviewers also concerned the lack of significant mathematics in the easiest items, those intended for the extensive number of low ability 15-year-old students around the world. Part of the test design for PISA 2012 was to include alternative test content to suit countries known or expected to be performing at a level markedly below the OECD average. Countries were able to choose two relatively easier clusters in place of two standard clusters, in order to provide better measures and richer descriptions of performances in the lower part of the PISA proficiency scale. Test developers had to write enough 'easy' items for this: if the given items did not cover the actual range of student ability then the test was not going to be maximally informative for the education authorities in a country. Marciniak in Chap. 5 of this volume discusses the confusion of significant and difficult mathematics in the PISA context. In fact, the statistical review showed that it had proved difficult to develop a large number of items suitable for these easy clusters, and the final choice of items for those clusters was from a smaller pool than had originally been hoped. After the field trial, all of the well performing very easy items went into the main survey, and more could have been used had they been available. On the other hand, the results of the field trial indicated that too many difficult items had been trialled. It seemed that test developers and reviewers were too optimistic about the mathematical literacy of 15-year-olds.

Using all of the data described above in a long and complex task, the MEG approved 90 paper-based and 45 computer-based items for the main survey. This was a few more than the minimum number of items required to construct the main survey instruments, so the ACER team had some flexibility in balancing all the framework requirements across the whole item pool, and also in balancing requirements within each cluster of items. These clusters were then arranged in the rotated design in booklets and online forms. In the end 72 new paper-based items and 41 computer-based items were used in the PISA 2012 main survey. To enable the 2012 results to be accurately put on the same scale as for previous PISA surveys, the 2012 booklets also included three clusters of link items (36 secure items) that had been used in previous surveys.

Reflections

As mentioned at the beginning, test developers on PISA quickly learned about the complex and sophisticated processes of developing such assessment instruments. We, as ACER mathematics test developers and as mathematics educators, had a number of reflections after the challenging journey of writing many items and assisting in the test development process in PISA 2012. This journey, alongside the knowledge about the actual performance of the items in both the field trial and the main survey, was very illuminating.

First, there were some very positive reflections from seeing the overall scores, scanning the actual responses of students from some countries and seeing other responses through the coder query process described by Sułowska in Chap. 9 of this volume. In many cases there were students who were able to respond and answer in a very sophisticated way, often showing unanticipated mathematical and real-world knowledge and insights. These students demonstrated much higher levels of mathematical understanding and knowledge than expected of most 15-year-olds, or alternatively a high ability to connect the mathematical content to the context—the ability to mathematically formulate problems from the real world, or to interpret mathematics in relation to the real world. The examples were very heartening to observe and it was an endorsement of the value and purpose of mathematical literacy.

In relation to CBAM, the test developers realised that this was only the starting point for computer-based assessment of mathematics. In the PISA 2012 survey, limitations were imposed by the time available for item development and by the information technology capacity in schools around the world, but also by the expectations of what computer-based tools 15-year-olds around the world would be able to manage at that point in time. There were a number of positives. One was the capacity to develop some highly interactive items that used combinations of animations and provided automatic calculations 'behind the scenes' to enable assessment of different and potentially deeper mathematical skills and understanding. As well, there was also the capacity to assess spatial and visual interactivity in a way not possible in a paper-based assessment. Some CBAM items definitely assessed skills that could not be assessed otherwise, hence broadening PISA's assessment of mathematical literacy. It was also heartening to observe students in cognitive laboratories being highly engaged with the CBAM tasks, and undertaking tasks with a very positive attitude and tending to persevere much more with them than with some of the paper-based items. It is hoped that future PISA surveys will extend the CBAM approach and feature more sophisticated, interactive computer-based mathematical literacy items.

Another key reflection is the observation, as mentioned earlier, that the spread of items written by seven professional test development centres across the globe significantly overestimated the mathematical literacy abilities of 15-year-olds around the world. The external reviewers of the 2012 PISA pool of items similarly over-estimated the expected mathematical knowledge of 15-year-olds. The psychometric analysis of the field trial data demonstrated this quite clearly—there were too few easy items and too many difficult items. This can also be interpreted as demonstrating that 15-year-olds around the world are not being well prepared in mathematics classrooms with the skills and knowledge to solve mathematical problems set within a real-world context. This is a challenge for education systems as we move further into the twenty-first century.

References

Australian Council for Educational Research (ACER). (2012). PISA: *Examples of computer-based items*. http://cbasq.acer.edu.au. Accessed 14 Nov 2013.

Forgione, K., & Saxby, M. (2011). *External validation of PISA 2012 mathematics field trial test item pool*. Unpublished meeting document of Mathematics Expert Group.

Ginsburg, H., Kossan, N., Schwartz, R., & Swanson, D. (1983). Protocol methods in research on mathematical thinking. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 7–47). New York: Academic.

Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. doi: 10.1787/9789264190511-en

Organisation for Economic Co-operation and Development (OECD). (2013b). *PISA 2012 released mathematics items*. Paris: OECD Publishing. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf. Accessed 8 Oct 2013

Chapter 8
Computer-Based Assessment of Mathematics in PISA 2012

Caroline Bardini

Abstract In 2012, when mathematics was again the major subject assessed, PISA included optional computer-based mathematics units for the very first time. This chapter will provide an overview of some of the key features of the computer-based units of PISA 2012 by addressing the following questions. What choices underpinned the design of the PISA units to be presented—and responded to—on a computer? What technological tools were available? Finally, what potential does a computer-based environment offer when it comes to assessing mathematical literacy and what are its limitations? These questions will be tackled taking into account the mathematical content knowledge, competencies and processes assessed as defined in the PISA 2012 Mathematics Framework.

Introduction

When calculators first made their appearance in mathematics classrooms, an avalanche of questions followed. Will students still be able to calculate? Will they lose their pen-and-paper skills? Will students still be able to *do maths*? Despite the many research studies that clearly show benefits in learning mathematics when calculators are appropriately used in the classroom, the debate is still lively and far from being closed (see for example the National Council Teachers of Mathematics summary by Ronau et al. (2011)).

A similar scepticism rekindles discussions that arose decades ago with the now growing availability of computers to students. Abundant research that focused in particular on the question of impact of different software—both commonly used desktop applications and software specifically designed for the teaching and learning of mathematics—on students' learning and understanding of mathematics has flourished ever since. And when it comes to using technological tools in assessment, the subject is particularly sensitive.

C. Bardini (✉)
Melbourne Graduate School of Education, The University of Melbourne,
234 Queensberry St, Melbourne, VIC 3010, Australia
e-mail: cbardini@unimelb.edu.au

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_8

It is not my aim to add to the pile of papers that make up the above debate, as I believe that the main character of the discussion oftentimes misses the real point. This is not about 'whether or not' to incorporate computers in the learning of mathematics (and this includes assessment), rather it is about 'how to do so'. It is undeniable that computers are nowadays part of everyday life and that they are of significant importance in the workplace. Burying one's head like an ostrich would only deprive us of appreciating the twenty-first century landscape with all its potentialities.

In the PISA 2012 Mathematics Framework (see Chap. 1 by Stacey and Turner in this volume), incorporating computers in mathematics assessments appears as an obvious fact: "a level of competency in mathematical literacy in the twenty-first century includes usage of computers" (OECD 2013, p. 43). Hence, following 2006 when PISA implemented computer-based science assessment, and after 2009 when it included an optional digital reading assessment, 2012 marked another major innovation in PISA. 2012 was when PISA included for the very first time an optional computer-based item assessment of mathematics—the year when mathematics was again the major subject assessed. But what should one understand by 'computer-based assessment'? More specifically, what should one understand by 'computer-based' assessment of mathematics in PISA 2012? In other words, exactly what mathematics was assessed in such an environment? What technological tools were available? What choices underpinned the design of the PISA units to be presented—and responded to—on a computer? Finally, what potential does a computer-based environment offer when it comes to assessing mathematical literacy and what are its limitations? These are the questions I propose to tackle in this article, from the point of view of a mathematics educator, also a member of the Mathematics Expert Group for PISA 2012.

Assessing Mathematics with a Computer in PISA 2012

Computer-Based Assessment: Characteristics, Affordances and Challenges

Despite an apparent contradiction, the following clarification is crucial for understanding what lies behind the notion of PISA's 'computer-based assessment'. It is of utmost importance to acknowledge that this type of assessment is not just an 'assessment on computer'. The units—and students' responses—are certainly presented on computers, but this must be distinguished from what could be interpreted as 'an electronic version of a paper-based unit'. As trivial as this distinction may appear to be, it is worthwhile highlighting it. In fact, the process of designing a computer-based item is far from consisting of different disconnected stages, that is to say, it does not follow a pattern such as: one team designs a paper-and-pencil task, then hands out to a technical team who 'transfers' it into a computer. Although the computer-based items were indeed originally presented on paper (item writers do not necessarily have programming skills), those items were, from their very first versions, designed with the anticipation of the fact that a range of electronic tools were available. Obviously there was at the end the need for a technical team to program and implement such units into a computer environment, but the item writers did design the different tasks with the aim of making the best use of all potentialities the computer environment could offer.

It is also important to note that the idea of incorporating computers in PISA 2012 mathematical literacy assessment was not primarily driven, for example, by the desirability of automated marking of the responses—clearly attractive when it comes to rating hundreds of thousands students' answers from over 60 countries. Various reasons underpinned the choice for a computer-based assessment and these can be viewed as responding to two aspects of the rationale. The first one, mentioned earlier, relates to the recognition of the importance of computational tools in today's workplace:

> For employees at all levels of the workplace, there is now an interdependency between mathematical literacy and the use of computer technology, and the computer-based component of the PISA survey provides opportunities to explore this relationship. (OECD 2013, p. 43)

The second one relates to the potentialities offered by the computer environment:

> the computer provides a range of opportunities for designers to write test items that are more interactive, authentic and engaging. (Stacey and Wiliam 2013). These opportunities include the ability to design new item formats (e.g., drag-and-drop), to present students with real-world data (such as a large, sortable dataset), or to use colour and graphics to make the assessment more engaging. (OECD 2013, p. 43)

But the essence of incorporating a computer-based assessment goes far beyond engagement and motivation and constitutes the core of every such item: to assess mathematical literacy in a way otherwise not possible—or at least too onerous to be considered. This is specifically what makes the computer-based items far from 'electronically transposed pen-and-paper tasks' and it is precisely what constituted one of the many challenges of this major area of innovation for PISA 2012. Since they were not merely electronic versions of paper-based items, computer-based items were particularly challenging to design as they added to the already complex task of having to create mathematical units that follow the different features described in the PISA Framework (balance between the different mathematical content categories, context categories and processes assessed, ranges of difficulty, etc.), and also keep to a minimum the load arising from information and communications technology (ICT) demands of the item. This is clearly acknowledged in the PISA 2012 Framework (OECD 2013).

What Competencies Assessed, with What Tools?

There are basically two types of mathematical ‘competencies’ (as referred in OECD 2013 p. 44) that are assessed in the computer-based units: those that are not dependent on the specifics of the environment (pen-and-paper versus computer) and those, on the contrary, that “require knowledge of doing mathematics with the assistance of a computer or handheld device” (p. 44). The former mathematical competencies are exactly the same ones that pen-and-paper units assess and these are tested in *every* computer-based item. The latter are present in some items only and, as described in PISA 2012 Framework, include the following:

- Making a chart from data, including from a table of values, (e.g., pie chart, bar chart, line graph), using simple ‘wizards’
- Producing graphs of functions and using the graphs to answer questions about the functions
- Sorting information and planning efficient sorting strategies
- Using hand-held or on-screen calculators
- Using virtual instruments such as an on-screen ruler or protractor
- Transforming images using a dialog box or mouse to rotate, reflect, or translate the image.

Amongst the many challenges item developers were faced with (see Chap. 7 by Tout and Spithill in this volume) and especially because of (i) the innovative character of such tests in mathematics units for PISA and (ii) the very tight timeframe that separated all the item creation stages (original version, programming, implementation and trial) was the fact that none of the electronic tools used in computer-based items pre-existed. For licensing reasons in particular, no existing software or tool could be used and although the international contractors had previously developed the delivery systems for the computer-based units of both Science and Reading, these were not exported to Mathematics. The programmers had indeed to design from scratch and within a very limited timeframe a wide range of electronic tools that best opened opportunities for “computation, representation, visualisation, modification, exploration and experimentation on, of and with a large variety of mathematical objects, phenomena and processes” (OECD 2013, p. 43). It is hoped that, as developers and item writers get more familiar with the underlying principles of a computer-based assessment, future PISA administrations will present even richer and more sophisticated items. Indeed, as noted in the PISA 2012 Mathematics Framework (OECD 2013), “PISA 2012 represents only a starting point for the possibilities of the computer-based assessment of mathematics.” (p. 43). Having said that, despite the complexity of the task, developers nevertheless produced computer-based items of a considerable range of types and formats, which reflect the notion of mathematical literacy as defined in PISA. The next section will use some released items (ACER 2012) to illustrate and further analyse the different types of tools available in this optional assessment taking into account the mathematical content knowledge, competencies and processes assessed.

Computer-Based Assessment of Mathematical Literacy in PISA 2012: Some Examples

Basic Tools and Features

As stated in the previous section, an on-screen calculator similar to pocket calculators or those present in commonly used desktop applications and mobile phones was available in every item. It included the four basic operations and square root, and was able to be customised and offered in different versions according to the standard notation (e.g. for division) of each participating country (see Fig. 8.1). This is just one example of how translation of items for use around the world requires attention to mathematical and format issues as well as the expected linguistic issues, as is described by Turner in Chap. 6.

As was permitted in the paper-based survey, calculators (real and virtual) were available not only because they are in some countries normally used in schools (hence potentially providing informative comparison of students' performance across different education systems) but also because assessing mathematical literacy goes beyond assessing computational skills—note that in many cases, numbers involved in items are carefully chosen so to encourage and ease eventual mental computations. The availability of the tool potentially relieves the burden of computation and helps students focus on the higher order mathematical thinking required by the task.

Amongst the most basic—yet important—features of computer-based units are the ones related to students' engagement and motivation. At their lower level of sophistication, one can name colourful presentations, three-dimensional representation of objects that can be rotated, moving stimulus, etc. Interactivity is also part of the basic features that a computer environment offers, but even at its most basic form (e.g. an online calculator) it can be an important asset when it comes to trying to assess aspects of mathematical literacy that would otherwise be too onerous either for students or for coders. Figure 8.2 provides an example of a more interactive item.

Fig. 8.1 Three versions of the on-screen calculator according to countries' standards (multiple combinations possible)

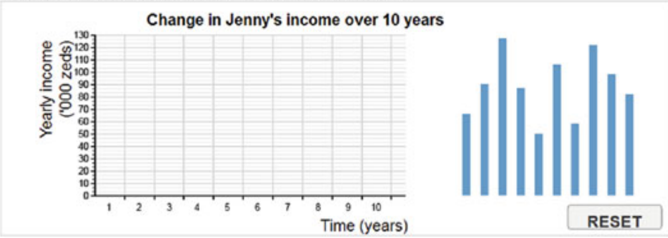
GRAPHS (format changed from original)

Jenny has worked at the same company for 10 years. Each year she recorded her yearly income. Her yearly income includes her annual salary plus any extra cash payments.

Jenny's annual salary **increased** by the **same** amount each year.

In years 4 and 9, however, she received an extra cash payment. These two payments were both of the same value.

The diagram below shows a set of labeled axes. The bars on the right of the diagram represent Jenny's ten yearly incomes but are not in the correct order.



Question 3: GRAPHS *CM010Q03*

Drag and position each of the bars onto the Time (years) axis to show how Jenny's yearly income changed over the 10-year period.

Fig. 8.2 CBAM item CM010Q03 Graphs Question 3 with drag-and-drop functionality (ACER 2012)

CM010Q03 Graphs Question 3, set in the *Scientific* context category, assesses mathematics from the *Uncertainty and data* content category and the *Employ* mathematical process. Students are required to order bars on a graph so that they are consistent with the given information. Full credit is given when all ten bars representing Jenny's income are correctly placed on the graph. The correct placement shows the bars in increasing order except for her income in years 4 and 9 where extra cash payments were made. This item was of above average difficulty in the field trial, with only 11 % correct. The response time of 80 % of students was less than 167 s. It had relatively low discrimination in 9 field trial countries, and therefore it was not used in the main survey.

The drag and drop functionality available to students works both ways: from right to left (group of bars to diagram) and conversely. Note that when dropping from right to left, bars are automatically centred on the corresponding intervals and their bases are positioned exactly along the time axis (which enables a reading of the yearly income to be independent of the precision of students' drag-and-drop action). This feature also enables dragging bars next to others, allowing students to, for example, easily compare heights before positioning the bars on the diagram. With such characteristics, the drag-and-drop functionality along with the possibility of having multiple attempts (reset button) allows students to focus on the mathematical features of the item (understanding a constant increase in value, interpreting graphically the extra cash received, etc.), instead of having to concentrate their efforts on drawing skills and precision, which are not targeted by this item.

Also, it seems hard to imagine a meaningful equivalent paper-based item. We could alternatively have a similar setting that displays an empty diagram with labelled graph axes on the left and a group of bars on the right. Bars could possibly be labelled and students could then be asked to write down the appropriate sequence of bars (which would thus deprive them of experiencing the graphical interpretation/representation of the evolution of the income over the years—the meaningfulness of such a task is hence questionable) or students could be asked to draw them on the empty graph (closer to the task set in the computer). In either case, the reading of the height of each bar is a potential initial problem. If originally displayed on a blank background without any grid as in the computer-based version, it would require, depending on students' strategies, a fastidious process and/or an additional drawing accuracy to determine the constant yearly increment of income (key to finding the appropriate answer). Even if students realise that one can begin by only comparing the two smallest bars to find the constant increment, to be accurate, this increment would have to be compared with the difference of height between—at least—another pair of bars with adjacent heights. The value of the heights (or eventually the increment between them) might vary according to accuracy of either (i) measuring the actual height of the bars with a ruler (which then would require a further conversion into the graph's scale) or (ii) transporting the heights into the diagram (by drawing a line parallel to the time axis, provided that the base of bars is aligned with the axis). Students might alternatively or subsequently perceive the need to find out the height of all bars before embarking on drawing of the diagram, which could turn out to be quite painstaking.

Another possible—and maybe more likely—paper-based version of this item could take this form: given the value of the first two incomes (or any two consecutive pairs of incomes excluding year 4 and 9) or their equivalent bars already drawn on the diagram, ask students to complete the graph according to the stimulus information. It is easy to see that the values of the income for years 4 and 9 become an issue, unlike on the computer-based version. In a paper-based item, one would have to either specify the extra cash or explicitly inform students that they should arbitrarily choose the amount. One of the benefits of the computer-based item is the fact that it is up to students to figure out that the exact value of the extra income is not relevant for solving the problem.

Many scenarios for a pen-and-paper version of item shown in Fig. 8.2 can be conceived. It is not our aim to provide an exhaustive range but this quick glance at some possibilities clearly highlights the benefits for using a computer environment, and the great potential for introducing substantially changed cognitive demands depending on what item design choices are made. Not only does the version used here emphasise students' mathematical thinking, but the task itself seems to be less artificial in its set-up. The interactivity feature of a computer-based assessment, which could have been perceived as a superfluous tool, can, when appropriately designed, become a powerful feature of high relevance for assessing mathematical literacy.

A Wide Range of Opportunities

Interactivity to Support Mathematical Thinking

Other than the drag-and-drop functionality, which can be seen as amongst the most basic types of interactivity when it comes to supporting students' mathematical thinking, interactivity can appear at a more advanced level, especially when designed to target competencies such as "sorting information and planning efficient sorting strategies" as listed in the Framework (OECD 2013, p. 44). Figure 8.3 provides an example. The unit CM038 Body mass index consists of three items, requiring students to derive information from a partially functioning website. Although the website has been specially constructed for the item and students doing the assessment are not connected to the internet, the website is authentic in the sense that there are many websites like this.

CM038Q03 Body mass index Question 1 involves the *Uncertainty and data* content category, and the *Interpret* process (make inferences from a set of graphs), within a *Societal* context. This is another example where a computer-based version of a task supports a strong focus on the mathematics being assessed. The website is partially functioning in the sense that students can click on the buttons to show or hide any of the six graphs. By default, all the six graphs are displayed, but not all of them are required to answer the true/false statements.

The truth value of the second statement is possibly easier to determine than the first one, as it explicitly indicates which value is relevant to answer the question, namely the lowest 5 % BMI (for both boys and girls). Although not essential, it is

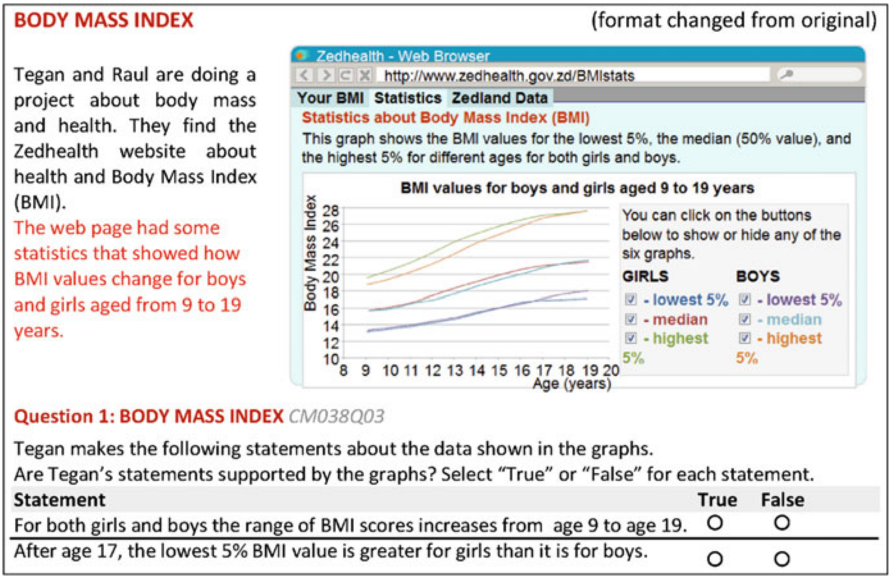


Fig. 8.3 CBAM item CM038Q03 Body mass index Question 1 (ACER 2012)

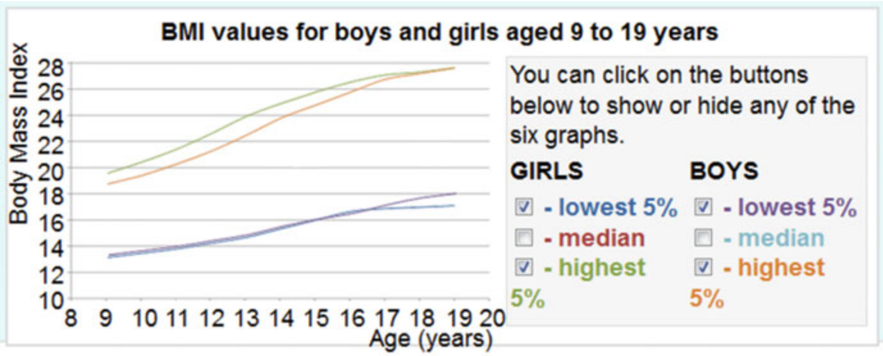


Fig. 8.4 Screenshot of CM038Q03 Body mass index Question 1 graphic without 'median' (ACER 2012)

expected from the way in which the instruction is worded ("You can click on the buttons below to show or hide any of the six graphs.") that students select the corresponding two graphs for the lowest 5 % BMI for boys and girls and hide the four others in order to answer the question (these two graphs are however sufficiently close one to another and separated from the other graphs for the latter not to be a visual distraction). One can think that a similar question on pen-and-paper could be envisaged; this issue will be discussed later on.

The first statement is, on the other hand, less obvious to decipher and the question ultimately requires students to adopt an efficient sorting strategy. The statement is indeed rich in information and along with their strategic skills, students will have to demonstrate appropriate understanding of diverse mathematical (statistical) concepts. One of the key issues students are faced with is the notion of 'range of BMI scores' and more precisely its translation into the graphical register. In other words, students will have to select which one(s) of the three given different BMI values (lowest 5 %, median—which definition is recalled—and highest 5 %) is (are) relevant to answer the question. In particular, acknowledging that the median value provides superfluous information (and hence that the two corresponding graphs can be hidden) is essential. Once this is discerned, students would probably deselect the median graphs for boys and girls, obtaining the screen as shown in Fig. 8.4.

Another key step is to understand and graphically interpret what it means for a range of values—in this case BMI scores—to increase and, at the same time, what is meant by 'for both boys and girls'. The specifics of the actual graphs reinforce the idea that, in order to judge the truth of an 'and' statement, one has to consider the two components separately. Since the graphs for boys and girls of both the 5 % lowest and 5 % highest values almost—and sometimes do—overlap, the need to display the pairs of graphs separately becomes indeed more evident. As contradictory as it may seem, analysis of the statement 'for both boys and girls' requires the students precisely to *not* display both sets of graphs. The interactive feature of this item to show or hide graphics goes beyond the added-value of students' engagement and motivation: it promotes substantial mathematical thinking, and allows the assessment of key mathematical knowledge.

It is easy to see why a pen-and-paper version of this item may not be as appropriate or rich as this computer-based item. It is also worth noting that, in order to add authenticity, the unit has been set as simulating a web-site, with two of the three tabs ('Your BMI', 'Statistics' and 'Zedland data') used to support the different items. The usage of a computer environment to replicate the usage of a computer in real life will be further discussed in the last paragraph of this section. Although PISA items often use real data sourced from a particular country, PISA items nearly always replace specific country information so that cultural biases are avoided. Instead the fictitious country Zedland with its currency the zed is used.

Geometrical Tools: Support for Students' Work as Well as for Coding Responses

Amongst the richest electronic tools to support students' mathematical thinking, including conjecturing, generalising and proving skills, are the various dynamic geometry packages nowadays commonly used in mathematics classrooms. Given the item developers' tight schedule, replicating such a complex tool with all the usual features was certainly not feasible. However different key features of dynamic geometry packages were incorporated in the computer-based units. These included being able to construct and/or rotate two- or three-dimensional shapes and objects, use virtual rulers to measure distances, dynamically change the shape of given two-dimensional figures, create points or lines on shapes, etc. These features, such as the one shown in Fig. 8.5 are of particular value when it comes to encouraging students' investigations in their search for, for example, specific geometrical properties of shapes.

It is not unusual to see tasks that aim at exploring the relationship between area and perimeter of given shapes in secondary—and in some cases even primary—mathematics classes. Setting tasks of optimisation such as CM012Q03 Fences Question 2 shown in Fig. 8.5 (maximal area with minimal perimeter—specifically relevant for the given *Occupational* context) on a computer offers a particularly mathematically rich environment as it potentially helps students to actively experience variation, often acknowledged as a stimulus for learning and awareness (Marton and Booth 1997) as well as for gaining mathematical knowledge (Watson and Mason 2005; Leung 2008). In fact, CM012Q03 Fences Question 2 simultaneously displays a geometrical representation of given shapes (rectangle and circle) and a table that records the corresponding values of their different features (length, width, area, etc.) that is automatically populated whenever there is a change in the shapes (through a dragging action). This multiple representation allows students to better recognise the effect of the change of each of the shape's dimensions on their area and perimeter and separate out patterns with respect to fixed conditions. This tool supports students to draw inferences from specific instances and conjecture on

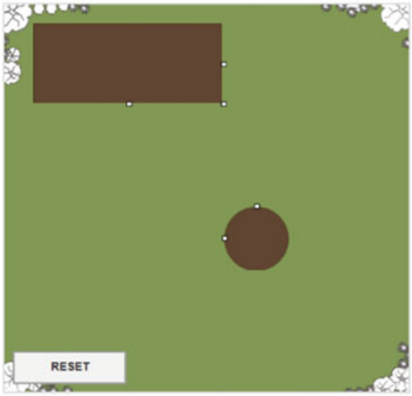
FENCES (format changed from original)

You can change the size of the garden bed by dragging a white handle (little white square).

The tables show the values of each of the measurements as you change the shapes.

Circle	
12	Diameter
38	Length of fencing used
113	Area

Rectangle	
35	Length
15	Width
100	Length of fencing used
525	Area



Question 2: FENCES CM012Q03

Are the following statements about the length of fencing and the shape and the area of the garden bed true or false? Select “True” or “False” for each statement.

Statement	True	False
If Linda uses a fixed length of fencing, then the maximum area you can obtain for rectangular gardens is when the shape is a square.	O	O
If Linda uses the same length of fencing, then a circle shaped garden bed gives a smaller area compared to a square garden bed.	O	O

Fig. 8.5 CBAM item CM012Q03 Fences Question 2 (ACER 2012)

the validity of a general case, which can support potential further algebraic work when seeking to make generalisations. This item was of average difficulty in the field trial, and 80 % of students answered within 115 s. It had low discrimination in 16 countries and was not used in the main survey.

Figure 8.6 shows another usage of interactive geometrical tools. By allowing students to create points and lines on the figures, the mathematical notion of star domain (in this unit tackled through the notion of star point) takes on a concrete dimension and its relevance in everyday life (e.g. for surveillance) is put forward in the last question of the unit, where more substantial mathematical thinking is required. Consider the item CM020Q01 Star points Question 1 shown in Fig. 8.6. At the same time that the computer setting of such task allows a more flexible assessment of students’ mathematical competencies (various correct answers are possible), it permits—and to some extent compels—a very precise marking scheme. In fact, the full credit code had to be devised by envisaging *all* possible answers as shown in Fig. 8.7. For shape 3 this includes any point in the lightly shaded triangular area; for shape 4 this includes any point *not* in the central square. This item was coded by computer. CM020Q01 Star points Question 1 was a relatively difficult item in the field trial. There were 11 % of students correctly indicating a star point for both Shapes 3 and 4, and 32 % correct for one shape. Only 11 % of students had missing responses. The item was completed by 80 % of

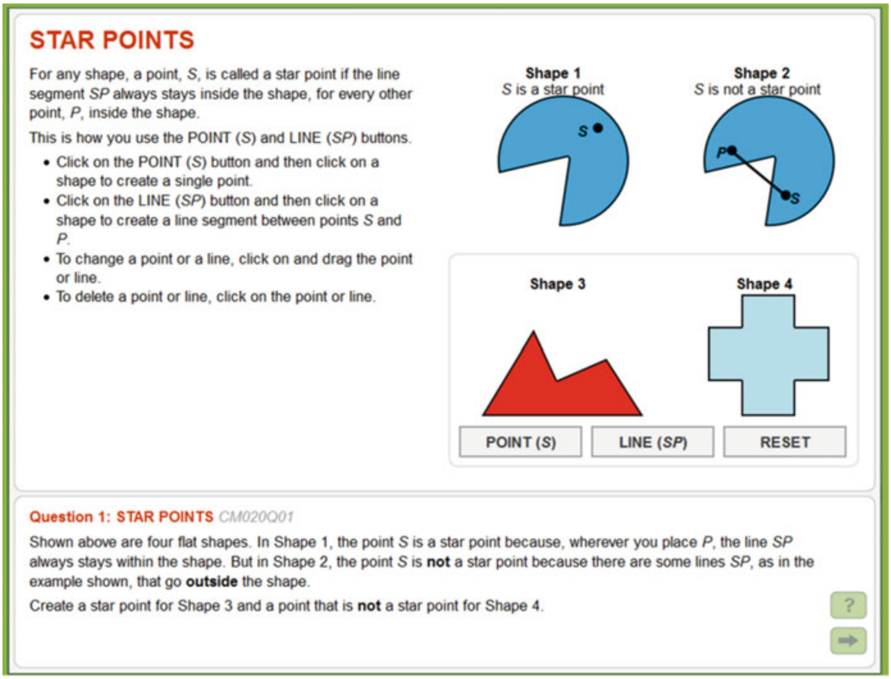


Fig. 8.6 Screenshot of item CM020Q01 Star points Question 1 (ACER 2012)

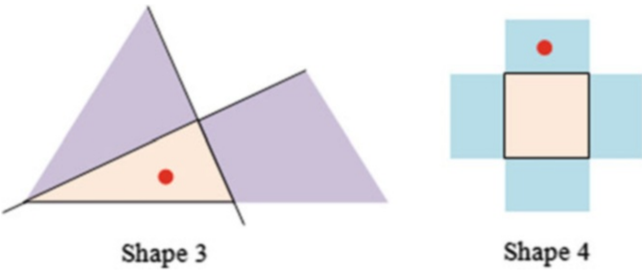


Fig. 8.7 Hot spots indicating regions of correct answers for CM020Q01 Star points Question 1 (ACER 2012)

students in less than 181 s. The unit was used in the main survey. A later item in the unit applied the star point idea to positioning of surveillance cameras.

Adding Authenticity: When Real-World Becomes Truly Real

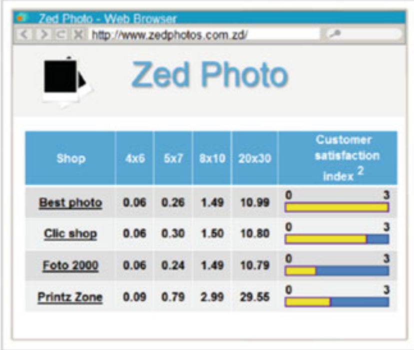
Developing authentic items has always been a major concern in PISA. In pen-and-paper units, authenticity is mainly conveyed through the context in which the item

PHOTO PRINTING (format changed from original)

The table shows the prices from four online digital photo shops. All the prices are in zeds per one photo.
Discounts and delivery charges[1] **are not included** in this table.

[1] For complete information, please click on the name of the shop in the first column.

[2] The satisfaction index is based on a survey of customers where they are asked to rate the quality of the service from 0 (lowest) up to 3 (highest). By hovering the cursor over each yellow bar, you can see the number of customers who answered the survey.



Shop	4x6	5x7	8x10	20x30	Customer satisfaction index[2]
Best photo	0.06	0.26	1.49	10.99	0 — 3
Clic shop	0.06	0.30	1.50	10.80	0 — 3
Foto 2000	0.06	0.24	1.49	10.79	0 — 3
Printz Zone	0.09	0.79	2.99	29.55	0 — 3

Question 1: PHOTO PRINTING CM030Q01

From the information in the table, is *Foto 2000* the cheapest for printing one photo of each format? Explain your answer.

Fig. 8.8 CBAM item CM030Q01 Photo printing Question 1 replicating a web page (ACER 2012)

is set. While this remains true for computer-based components as attested by the items discussed above (designing a garden bed, positioning security cameras in a shopping centre, etc.), setting the task on a computer environment provides the opportunity to explore the authenticity feature at an even higher level.

The previously analysed unit CM038 Body mass index shown in Fig. 8.3 provides one example where a fictitious web-page is presented. Although one can acknowledge that this potentially adds authenticity to the question, it does not make the most of what a computer environment may offer. The unit CM030 Photo Printing illustrates a further exploitation of such a feature. Figure 8.8 shows the initial screen of the unit.

As for CM038 Body mass index, this unit simulates an online activity—here comparing prices of printed photos for different online shops, but different features of such item contribute to adding authenticity to the task. The first one is that it includes information nowadays often present—or at least sought after—when shopping on line, namely the rating of the service or product by previous costumers. This is represented by horizontal bars showing scores from 0 to 3. Another more advanced feature is shown in Figs. 8.9 and 8.10.

CM030 Photo printing takes a step further on the simulation of a web-page, as it enables the user to interact with the content displayed on the screen. By clicking on any of the shop names or hovering over them with the mouse, students are able to gather further information to help them make appropriate judgements when comparing prices between the different shops. It is worth mentioning that no question

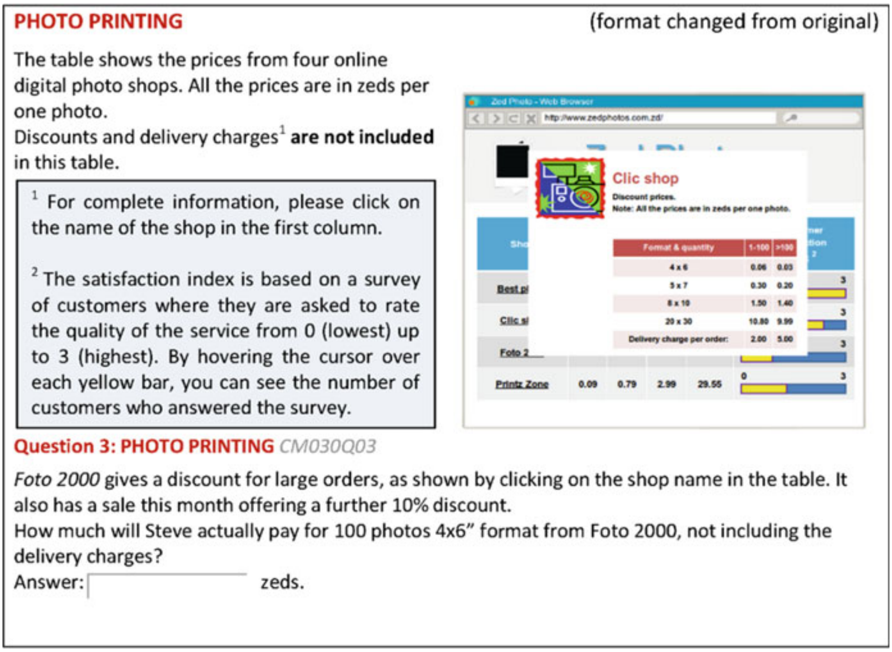


PHOTO PRINTING (format changed from original)

The table shows the prices from four online digital photo shops. All the prices are in zeds per one photo.
Discounts and delivery charges[1] **are not included** in this table.

[1] For complete information, please click on the name of the shop in the first column.

[2] The satisfaction index is based on a survey of customers where they are asked to rate the quality of the service from 0 (lowest) up to 3 (highest). By hovering the cursor over each yellow bar, you can see the number of customers who answered the survey.

Question 3: PHOTO PRINTING CM030Q03

Foto 2000 gives a discount for large orders, as shown by clicking on the shop name in the table. It also has a sale this month offering a further 10% discount.
How much will Steve actually pay for 100 photos 4x6" format from Foto 2000, not including the delivery charges?
Answer: ________ zeds.

Fig. 8.9 CM030Q03 Photo printing. Clicking on or hovering the mouse for further information (ACER 2012)

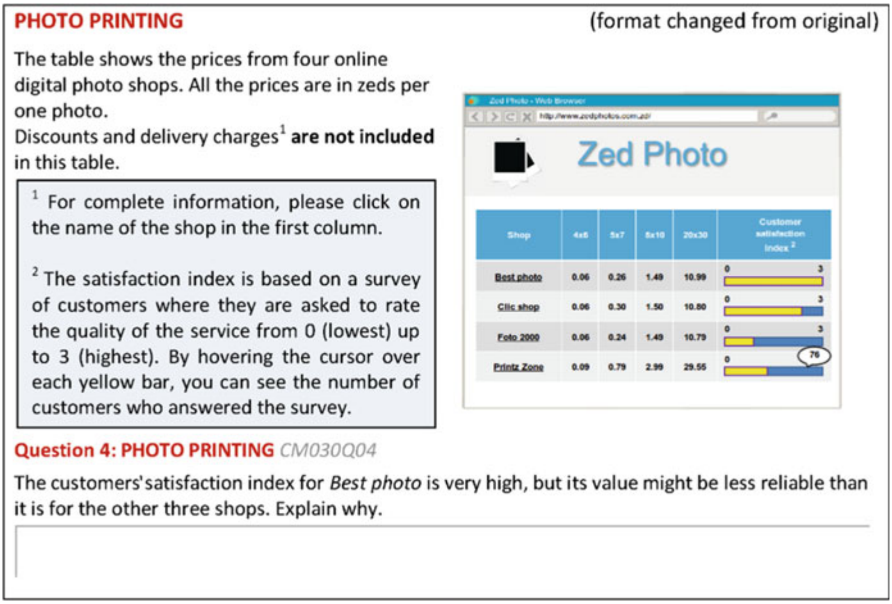


PHOTO PRINTING (format changed from original)

The table shows the prices from four online digital photo shops. All the prices are in zeds per one photo.
Discounts and delivery charges[1] **are not included** in this table.

[1] For complete information, please click on the name of the shop in the first column.

[2] The satisfaction index is based on a survey of customers where they are asked to rate the quality of the service from 0 (lowest) up to 3 (highest). By hovering the cursor over each yellow bar, you can see the number of customers who answered the survey.

Question 4: PHOTO PRINTING CM030Q04

The customers' satisfaction index for *Best photo* is very high, but its value might be less reliable than it is for the other three shops. Explain why.

Fig. 8.10 CBAM item CM030Q04 Photo printing Question 4 showing hovering the mouse for further information (ACER 2012) requires students to gather information from all four shops. In fact, it is only in Question 3 shown in Fig. 8.9 that students have to collate additional data on shop 'Foto 2000'. However, to add authenticity as well as assess students' ability to identify relevant information, further information is available for all the four shops.

Figure 8.10 shows another instance where there is a need to find further information (by hovering the mouse). Indeed, in order for students to analyse the reliability of the rating for 'Best photo' compared to the other shops, students need to understand the importance of the sample size that the given score is based on. By hovering the mouse over each rating bar, students are able to see the number of customers who have actually answered the satisfaction question and hence better compare the reliability of the different shops' scores. And apart from adding authenticity to the task, the functionality illustrated in Figs. 8.9 and 8.10 also substantially relieves the *communication* demand of the unit (see Chap. 4 of this volume), which would require lengthy text in a paper-based item.

Concluding Remarks and Perspectives

This chapter presented and analysed some of the features of PISA 2012 computer-based mathematics items as theoretically described in the Framework (see Chap. 1 of this volume) and as implemented in practice. Although limited to examination of the publicly released items, these already illustrate a range of characteristics. Even this first implementation of CBAM demonstrates an array of potential which I hope will be further exploited in future PISA surveys. Indeed, this chapter has shown that although setting the test in a computer environment may increase students' engagement and motivation, the reasons for integrating such components in a mathematical literacy assessment go far beyond these and actually do provide opportunities to give a more rounded picture of mathematical literacy.

Just as "PISA 2012 represents only a starting point for the possibilities of the computer-based assessment of mathematics" (OECD 2013, p. 43) the present chapter is only a partial discussion on this matter. Many avenues that have not been explored here are worth considering. Amongst these, there is the analysis of the whole range of extra information that testing on a computer allows to be gathered that could supplement and refine the analysis of students' responses. These include tracking students' clicks. Has the student repeatedly selected and deselected boxes, suggesting some hesitation? Has the student clicked on the relevant tools or regions of the screen for a given question? In what order? Recording the time spent on each item could be used to modify the results of the survey (rather than only to arrange clusters of an appropriate number of items as was done for 2012 using field trial data) or the tools used. Has the student used the on-screen calculator? When? This additional information could be of particular help when it comes to, for example, examining students' responses to multiple-choice questions. Then there is the obvious question that keeps fuelling the debate whenever it comes to assessing on computer, namely the effect of this specific support on students' performances. As pointed out in the PISA 2012 Framework

> Research has been conducted on the impact a computer-based testing environment has on students' performance [original references omitted] and the PISA 2012 survey provides an opportunity to further this knowledge, particularly to inform development of future computer-based tests for 2015 and beyond. By design, not all computer-based items will use new item formats, which might be helpful in monitoring the (positive or negative) impact that new item formats have on performance. (OECD 2013 p. 43)

Specific studies that compare paper-based and computer-based modes of assessment on parallel items, particularly when it involves large-scale testing, have been previously conducted and PISA computer-based assessment of science has already been explored. Similar studies that would now focus on the mathematical literacy competencies would be worth exploring, especially if computer environments will progressively become the main means of assessing students' performance.

References

Australian Council for Educational Research (ACER). (2012). PISA: Examples of computer-based items. http://cbasq.acer.edu.au). Accessed 14 Nov 2013.

Leung, A. (2008). Dragging in a dynamic geometry environment through the lens of variation. *International Journal of Computers for Mathematical Learning, 13*, 135–157.

Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah: Lawrence Erlbaum.

Organisation for Economic Co-operation and Development (OECD). (2013). PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Ronau, R. N., Rakes, C. R., Bush, S. B., Driskell, S., Niess, M. L., & Pugalee, D. (2011). Using calculators for teaching and learning mathematics. In K. D. King (Ed.), *Technology research brief*. National Council of Teachers of Mathematics. http://www.nctm.org/uploadedFiles/Research_News_and_Advocacy/Research/Clips_and_Briefs/2011-Research_brief_18-calculator.pdf. Accessed 14 Nov 2013.

Stacey, K., & Wiliam, D. (2013). Technology and assessment in mathematics. In M. A. K. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 721–752). New York: Springer.

Watson, A., & Mason, J. (2005). *Mathematics as a constructive activity*. Mahwah: Lawrence Erlbaum.

Chapter 9
Coding Mathematics Items in the PISA Assessment

Agnieszka Sułowska

Abstract Coding of student responses is one of the most important, but also difficult processes of the PISA assessment. This chapter explains how this is done, without assuming any prior knowledge of PISA assessment. First the resources and procedures that are used in the course of the coding process are described: the coding guide and the general principles of coding. For better understanding, actual items are used to illustrate the multilayer structure of codes. There is an explanation of the elaborate preparations for the coding process, both within a participating country as well as globally, aimed at reaching a common understanding of codes within the international community of coders. After the long period of careful preparations the actual coding takes place. The actual coding process is explained by sharing the author's experiences as a supervisor of coders during four consecutive PISA survey administrations. Examples illustrate the inevitable coding dilemmas, proving again and again that our students' creativity exceeds the imagination of the most experienced coding guide authors. Examples also show how PISA can resolve such dilemmas in a systemic way.

Introduction

Coding of student responses is one of the most important, but also difficult processes of the PISA assessment. Students have just completed the booklets and now, across the globe, their responses have to be transformed into codes in the most uniform way.

Why is this called coding, rather than marking or grading? There is a fundamental difference between PISA coding and the marking of students' papers, as practised by thousands of teachers on a daily basis: it has a different objective. When a teacher marks a student's work or a presentation, he or she really creates feedback information for the student. It is aimed at helping the student to recognise

A. Sułowska (✉)
Mathematics Section, Educational Research Institute IBE, 8 Gorczewska Str, 01-180, Warsaw, Poland
e-mail: a.sulowska@ibe.edu.pl

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_9

his or her strong and weak areas within the assessment scope, as the first step towards improvement.

In contrast, the over-riding objective of coding in PISA-like surveys is only to obtain the data from which a measure of mathematical literacy can be derived and applied to specified groups (countries, girls, boys, etc.). Also the coding needs to be carried out in many countries by many different people and in many different languages, so it must be as simple and robust and also as economical as possible. It is of utmost importance to get consistency across all these different groups so that differences in the measure of mathematical literacy reflect as nearly as possible differences in the students, and not systematic differences in how the assessors in each country have valued different responses.

This chapter aims to explain how those crucial issues are addressed in the PISA survey. Thus, it describes the resources and procedures that are used in the course of the coding process: the coding guide and the general principles of coding. For better understanding, actual items are used to illustrate the multilayer structure of codes. There is also an explanation of the elaborate process of preparations of the coding process, both within a participating country as well as globally, aimed at reaching a common understanding of codes within the international community of coders.

After the long period of careful preparation the actual coding takes place: stacks of booklets arrive at the coding venue, filled with the full richness of students' responses. Some of the items can be automatically coded, but this chapter is concerned with the items labelled in Chap. 7 as Constructed Response Expert and Constructed Response Manual, where expertise and judgement are required. The chapter explains the actual coding process, by sharing the author's experiences as a supervisor of coders during four consecutive PISA survey administrations. Examples illustrate the inevitable coding dilemmas, proving again and again that our students' creativity exceeds the imagination of the most experienced coding guide authors. Most importantly, examples show how PISA is prepared to resolve such dilemmas in a systemic way.

The Coding Guide Structure

Coding of PISA items involves assigning appropriate codes to students' responses. Codes available for each item are precisely described by the coding guide. The codes for each item are essentially defined at two or three levels: either *Full credit—No credit,* or *Full credit—Partial credit—No credit*. These descriptors were chosen to avoid formulations like: 'correct answer', 'partially correct answer' and 'incorrect answer'. It was done on purpose, to stress the fact that a *Full credit* code can be assigned to a solution that is not perfectly correct and also a *No credit* code can be assigned to a solution that is not completely wrong. The precise description of the level of accuracy of students' responses required for each code level is item-specific. In most items, the coding is only single digit, indicating full credit, partial credit, or no credit (2, 1, 0) in some items, and just full credit or no credit (1, 0) in others. The fact that some items have full credit coded as 2 and others have full credit coded as 1 does not indicate any weighting of the items in creating a total score. These codes are not totalled to get the students' results. Instead, the complex statistical processes used to calculate overall scores are based on Rasch-based item response modelling. They are described in OECD technical reports such as Adams and Wu (2003). As will be demonstrated below, some items have 'double digit' codes, which provide researchers with information about the solution processes that students use, but they do not change the allocation of full, partial or no credit.

Figure 9.1 provides an example of one item. PM977 DVD rental was a three-item unit, which was used in the PISA 2012 field trial then released (OECD 2013). Figure 9.1 shows the stimulus, Question 2, and the categorisation of this question and Fig. 9.2 shows the coding instructions for its double digit coding. The item was of above average difficulty. Eight different codes have been defined for this item: four at the full credit level (labelled with codes 21, 22, 23, 24 in Fig. 9.2) two at the partial credit level and two at the no credit level. Each code is defined by a description of the kind of student responses to which it will be applicable. In addition, most codes are illustrated by examples of actual students' responses, as displayed in Fig. 9.2.

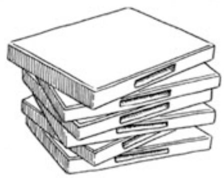
DVD RENTAL

PM977

Jenn works at a store that rents DVDs and computer games.

At this store the annual membership fee costs 10 zeds.

The DVD rental fee for members is lower than the fee for non-members, as shown in the following table:



Non-member rental fee for one DVD	Member rental fee for one DVD
3.20 zeds	2.50 zeds

Question 2: DVD RENTAL

PM977Q02 – 00 11 12 21 22 23 24 99

What is the minimum number of DVDs a member needs to rent so as to cover the cost of the membership fee? Show your work.

(space removed from image)

Number of DVDs: ....

QUESTION INTENT:

Description: Calculate and compare numbers in an everyday situation
Mathematical content area: Quantity
Context: Personal
Process: Formulate

Fig. 9.1 PM977Q02 DVD rental Question 2 with categorisation (OECD 2013)

Full Credit

Code 21: Answer 15 [*Algebraic solution with correct reasoning.*]

$3.20x = 2.50x + 10$
$0.70x = 10$
$x = 10 / 0.70 = 14.2$ approximately
but whole number solution is required: 15 DVDs

$3.20x > 2.50x + 10$
[Same steps as previous solution but worked as an inequality].

Code 22: Answer 15 [*Arithmetical solution with correct reasoning.*]

For a single DVD, a member saves 0.70 zeds. Because a member has already paid 10 zeds at the beginning, they should at least save this amount for the membership to be worthwhile. 10 / 0.70 = 14.2... So 15 DVDs.

15 x 3.2 – 10 = 38, 15 x 2.5 = 37.5. So 15 DVDs is cheaper for members.

Code 23: Answer 15 [*Solve correctly using systematic trial and error.*]

10 DVDs = 32 zeds non-members and 25 zeds + 10 zeds = 35 zeds for members. Therefore try a higher number than 10. 15 DVDs is 48 zeds for non-members and 37.50 + 10 = 47.50 zeds for members. Therefore try a smaller value: 14 DVDs = 44.80 zeds for non-members and 35 +10 = 45 zeds for members. Therefore 15 DVDs is the answer.

Code 24: Answer 15. Without reasoning or working.

Partial Credit

Code 11: A correct method (algebraic, arithmetical or trial and error) but minor error made leading to a plausible answer other than 15.

Code 12: Correct calculation but with incorrect rounding or no rounding to take into account context.

- 14
- 14.2
- 14.3
- 14.28 …

No Credit

Code 00: Other responses.

Code 99: Missing.

Fig. 9.2 PM977Q02 DVD rental Question 2 coding guide (OECD 2013)

As shown by this item, the relevant part of the coding guide is released with the item, although coding teams are given considerably more detailed instructions that are unpublished. Extracts from this unpublished material are used in this chapter.

All codes for this item are double digit codes. The first digit defines the code level (which is the score for the item, used to calculate performance); the second is specific for a group of responses at that level and reflects a method students used to approach the problem or a type of student error. In PM977 DVD rental Question 2, the full credit code level is 2, (so the associated codes are 21, 22, 23 and 24), the partial credit code level is 1 and the no credit code level is 0.

As is evident in the coding instructions in Fig. 9.2, a student's response can be assigned a code at the full credit level only if it contains the correct answer of 15 DVDs. A specific code from that level is selected according to the method applied by the student to obtain that answer. If the student just gave the correct number without any explanation and hence we cannot infer by which method the number was determined, code 24 was used.

At the partial credit level we have two codes. The first, code 11, was applied to responses in which a student had applied a correct method but also made an arithmetic error that resulted in a number of DVDs different from 15. The second code at this level, code 12, was used when the method and calculations were correct, but the final result was not rounded and hence the answer to the question is also not 15. Thus, the two codes make a distinction between a general error (code 11) and a specific mistake (code 12).

At the no credit level we also have two codes. The first, code 00: *Other responses* is applied to all responses not covered by the higher level codes. For example, when a student had used a correct method, but made an arithmetic error and also did not round the obtained real number then the response does not fit any of the criteria defined for the codes mentioned earlier and hence the response is coded as 00. This is an example of a response that is not completely wrong (a correct solution method was used), but it still gets a code from the no credit level. The same code is assigned to all completely wrong responses.

Code 00 covers also all responses that were first written down but later either rubbed out or crossed out by the student, whether legible or not. It is also assigned to all responses like 'it is too difficult', 'I do not have enough time', 'this is silly' or even when a student puts in the solution space a question mark or just any mark. In all such instances it is assumed that the student has read the item, but does not have the ability to provide a solution.

At the no credit level we also have code 99: *Missing*, which is applied when the solution space is completely empty and there are no signs indicating that the item was read by the student. There is one more special code 97: *Not applicable*. This is used when it was not possible for the student to answer the question for reasons independent of ability, for example when the print was not legible or an essential part of the supporting drawing or graph was missing in the student's booklet. A further code is also applied in the data analysis stage, after coding has been completed, because of its interest for research. This *Not reached* category is applied to all of the items in the booklet beyond the last one reached by the student, *i.e.*, when all the following items got code 99. There are likely to be different reasons for missing responses. For example when insufficient time may be a factor, missing responses may be interspersed with answered items, especially because the items are not arranged in order of difficulty. An uninterrupted sequence of missing responses at the end of student booklets is not included in the calculation of item difficulty parameters, but such responses are treated as 'incorrect' for the purpose of estimating student abilities.

General Rules of Coding

The basic tool used by the coder is, of course, the coding guide. However, despite the great attention paid to eliminating the subjectivity of coding, by means of very carefully formulated code descriptions and by selection of representative response examples, the coder has sometimes to make the decision how to classify a particular borderline response and hence to decide where the subtle borders between different codes are located.

To make such decisions coherent, several general rules are defined in the coding guide. The first fundamental and intuitively obvious rule is that spelling and grammar mistakes should be ignored. In PISA Mathematics, the assessment measures mathematical literacy—it is not a test of written expression. For the same reason also a student's arrangement of the response plays no role. For example, it does not matter when a student presents a descriptive solution instead of circling one of the words YES or NO, or when a student positions the response outside the expected response space (e.g., on the margin, next to the picture etc.)

The second rule states that when the student's response does not fit any code description, the coder should consider whether the student has understood the substance of the question and to what extent has demonstrated the ability to answer the question. Each code in the coding guide covers a certain class of responses, which correspond to a certain class of students' abilities. Some codes are defined by indicating typical students' errors, which—in turn—identify the lack of certain abilities. In the case of a response not fitting any code description—in most cases this is a partially correct solution—the coder must try to identify the reason for the student's mistake and make a judgement about the student's abilities. Next, the coder should compare these abilities with those associated with particular codes and then assign the code best fitting the response.

The third rule states that coders should avoid applying a deficit model. In other words, they should avoid deducting 'points' for anything that falls short of a perfect answer or for each error. This rule also gives the student the benefit of any doubt about the response when it seems reasonable to do so. For example, coders should be ready to accept a certain degree of informality or even a chaotic presentation of the solution. Also they should not penalise solutions employing mental arithmetic.

The fourth rule concerns responses that contain more information than required by the question or that is irrelevant to the question. The main task of the coder is then to consider whether or not the elements of the response contradict one another. If a contradiction occurs, the no credit code is applied. For example, if the expected answer is a number and a student provides two different numbers, without indicating (or crossing out) one of them, then code 0 is assigned even if one of the two numbers is correct. On the other hand, if no elements of the response contradict each other, the coder should ignore the irrelevant information and assign a code to the relevant part of the response.

Coding Preparation Process

International Coders' Training

Prior to each survey, both the field trial and the main survey, there are organised international meetings for persons supervising the coding process in the participating countries. At each of those meetings most of the time is devoted to joint coding of a selected set of solutions. Those solutions represent typical responses, illustrating the particular code categories, as well as problematic responses, not fitting directly any of the code descriptions. The process of coding those solutions is often accompanied by fierce discussion. This 1-week-long joint work results in a set of solutions with codes assigned. They enrich the set of example responses illustrating the particular code categories and can later help coders to make decisions in dubious cases. They are also used as a source of training materials for the national coders' training, which is held in each country.

National Coders' Training

The general rules of the national coders' training are defined by PISA procedures. To illustrate those rules and their implementation, let us review the coders' training process in Poland.

In all PISA survey administrations, we have decided to employ as PISA mathematics coders, students who are studying for a Masters or PhD degree in mathematics from the University of Warsaw or the Warsaw University of Technology. Each time we have found their work highly satisfactory. They have put every effort into fully understanding the coding guide and were truly devoted to applying it with full precision. Multiple coding statistics have each time confirmed a high degree of agreement of their codes. Also their sharp mathematical brains have helped to resolve the mysteries of many obscure responses.

Over the years, I have had quite a few meetings with Polish teachers of mathematics, presenting the PISA results to them. Many of them find it difficult to accept strict rules of coding, in full accordance with the coding guide. They were not able to distance themselves from their private rules and convictions concerning the evaluation and rewarding of individual student's work, which they had developed in their school practice, where teaching good mathematical practices is the main goal. In particular, they usually had very strong, although quite subjective, opinions about what it means for a solution to be correct. They found it difficult to accept that for measurement purposes, full credit could be assigned to a student's solution that is not perfect. Also, they were not flexible enough to accept the assignment of no credit to solutions that are completely wrong as well as to solutions that are partially correct, but not covered by higher codes. These observations made me very careful when recruiting and later training my PISA coders.

Before the start of the training, applicants for PISA coding positions have to study carefully a few of the released PISA items with the corresponding codes from the coding guide as well as the general rules of PISA coding. During a meeting, materials are thoroughly discussed and the participants are encouraged to ask any questions. After answering all questions and clearing up all their doubts, they are given the task of coding a dozen or so sample student responses to each of the discussed items. Those candidates who perform best are invited to the main coder training.

During the main coder training, the coders acquaint themselves with the actual coding guide that they will be using in the coding process and review once again the general principles of coding. At the training preceding each coding round, items that are about to be coded are discussed again. Next, a training round of coding occurs, based on the training materials prepared earlier. Students' responses included in the training materials originate both from the international coders' training as well as from students' response booklets from Poland. The trial coding consists of two sessions. During the first session, coders jointly assign codes to a first set of students' responses from the training materials and discuss their decisions. The aim of this session is to reach precise understanding of rules of the item coding. In the second session, each coder independently codes a second set of students' responses so that their coding accuracy can be assessed by the supervisor. Responses for which full conformity was not reached are discussed again. Then the actual coding starts. The work of coders whose coding during the training session did not fully comply with the expected results is carefully supervised.

Coder Query Service

It often happens during the actual coding that a coder has difficulty assigning a code to a student's response. Then they can ask the supervisor for help. This person, equipped with the experience of the joint coding at the international coders' training and also with the thorough knowledge of the coding guide enriched by a set of coded items, can help to make a decision. However, it can happen that the supervising person also has serious doubts concerning the code assignment to a particular solution. If the difficulty encountered concerns more than one case, a query can be sent to the coding department of the international contractors organising the PISA survey. A list of solutions causing coding difficulty received from different countries together with the correct codes assigned and supporting comments are distributed by the international contractors for use by all national coding teams. This document called Coder Queries provides an even bigger set of coding examples, which can be referred to in case of doubt.

Coding of a Sample Item

The unit PM978 Cable television was released (OECD 2013) after use in the field trial for the PISA 2012 survey but not in the main survey. The first question in the unit was multiple-choice so automatically scored and not dealt with by the coders. Question 2 is given in Fig. 9.3. In the field trial this question PM978Q02 was slightly easier than average.

As shown in Fig. 9.4, this item has a relatively simple, two level system of codes. At the no credit level we have the simplest possible set of codes: 00 and 99. At the full credit level there are also only two codes: 11 and 12. The distinction between codes 11 and 12 is quite clear. Code 11 is used when the student points out the general rule that the total number of households is essential information for interpreting the percentage. Code 12 is applied when the student just calculates quantities relevant to the problem. The three sample responses illustrating the codes are clear too—each of them quite extensively justifies the claim posed in the item.

An additional set of sample students' responses with codes assigned and extensive comments was assembled by the item development team for the international coders' training. This set contained, among others, the following responses:

Response 1	This is incorrect because France has a lot bigger population (24.5 million) whereas Norway only has a population of 2 million
Response 3	97 % of 24.5 million >97.2 % of 2 million
Response 5	The statement is incorrect because France has a much larger amount of households that own TVs
Response 7	The population of France is bigger than the population of Norway

Responses 1, 5 and 7 were assigned code 11. They fit the general code description, although they are less extensive than the sample responses quoted in the coding guide. The most laconic is response 7. Here the student mentions neither the percent calculation of the quantities being compared nor quotes any exact numbers showing the large difference of the population sizes. The rationale for assigning code 11 to this response (given in the unpublished documentation for coders) is as follows:

> We feel there is an implicit understanding of the percentages of cable TV subscribers (otherwise they would not have responded in the way they did), and that they recognised that the much higher total number of households owning TVs in France compared with Norway overrides the difference in percentage in Cable TV subscribers. So we are giving the student the benefit of the doubt that they had taken those percentages into account.

Response 3 received code 00, with the accompanying comment in the unpublished coding advice:

> The student has simply written down the numbers from the first two columns of the table—we feel if they were aware of the need to take into account the information in the last column they would have included those in their calculations (and it would have then been a clear code 12).

CABLE TELEVISION

PM978

The table below shows data about household ownership of televisions (TVs) for five countries.

It also shows the percentage of those households that own TVs and also subscribe to cable TV.

Country	Number of households that **own TVs**	Percentage of households that **own TVs** compared to **all households**	Percentage of households that **subscribe to cable television** compared to households that **own TVs**
Japan	48.0 million	99.8%	51.4%
France	24.5 million	97.0%	15.4%
Belgium	4.4 million	99.0%	91.7%
Switzerland	2.8 million	85.8%	98.0%
Norway	2.0 million	97.2%	42.7%

Source: ITU, World Telecommunication Indicators 2004/2005
ITU, World Telecommunication/ICT Development Report 2006

Question 2: CABLE TELEVISION *PM978Q02 – 00 11 12 99*

Kevin looks at the information in the table for France and Norway.

Kevin says: "Because the percentage of all households that own TVs is almost the same for both countries, Norway has more households that subscribe to cable TV."

Explain why this statement is incorrect. Give a reason for your answer.

(answer space omitted)

QUESTION INTENT:

Description: Understand proportionality based on data provided in a table
Mathematical content area: Uncertainty and data
Context: Societal
Process: Interpret

Fig. 9.3 PM978Q02 Cable television Question 2 with categorisation (OECD 2013)

Full Credit
Code 11: A response that says that Kevin needed to take into account the actual number of households with TVs for the two countries. [Accept "population" as a substitute for "households"].
He is wrong because there are over 22 million more households that own TVs in France, and even if only 15.4% subscribe to cable TV that is more than Norway.
Because the population of France is about 10 times more than Norway and there is only about 3 times as many households that subscribe to cable TV in Norway compared to France.
Code 12: A response that is based on calculation of the actual number of subscribers in the two countries.
Because France has 24.5 × 0.154 = approximately 3.8 million households that subscribe to cable TV, while Norway has 2.0 × 0.427 which is approximately 0.8 million households. France has more cable television subscribers.
No Credit
Code 00: Other responses.
Code 99: Missing.

Fig. 9.4 PM978Q02 Cable Television Question 2 coding guide (OECD 2013)

Among all the queries received by the international contractors, only four concerned the PM978 Cable television item. Two of them were similar to the responses 1, 5 and 7 from the coders' training, quoted above. One query (Query 5184) concerned the response: "There is a great difference in the number of households that own TVs in both countries." The student author of this response does not state precisely how large the difference between the population sizes of the two countries is, nor is it explicitly stated in which direction this difference works—to France or to Norway. This was even vaguer than the above responses, but it still was given code 11:

> The response implies understanding of the percentages of cable TV subscribers and recognition that the much higher total number of households owning TVs in France compared with Norway overrides the difference in percentage in Cable TV subscribers. So we recommend giving the student the benefit of the doubt that they had taken those percentages into account.

Another query (Query 5017) concerned a response that contained a small calculation error: "This is incorrect because France has a 23.5 million difference in the number of households that own a TV." In fact, the difference is 22.5 million, not 23.5 million. The coder submitted a query to the international contractors asking whether this student response was more like Responses 1 and 5 above, than Response 3. Again the decision was to give code 11. "We feel that the student response provided is most similar to [Responses 1 and 5 above] and should be scored accordingly as Code 11."

The Polish Experience of Coding the Cable Television Item

While coding over 1,000 Polish students' responses, coders came across several answers that were hard to code. Let us look at four examples:

Response P1	The number of households that own TVs is smaller
Response P2	Despite the fact that in Norway the percentage of households that subscribe to cable TV is about 3 times larger, the number of those households is about 12 times smaller
Response P3	Because more people live in France than in Norway and not everybody subscribes to cable TV
Response P4	France has over 15 % cable TV subscribers. But even if there were only 10 %, it would still amount to about 2.5 million subscribers. Norway has about 42 %, but even 50 % would give only 1 million subscribers. Hence France has more subscribers

Response P1 is close to the general description of code 11 in the coding guide shown in Fig. 9.4. However, it is far more terse than any of the examples provided there—the student did not use any numbers and did not indicate which country has fewer households with cable TV. For that reason coders had doubts whether such a general response deserves code 11. After comparing it with Query 5184 above, they decided to assign code 11.

Response P2 above is very close to the second example for code 11 in the coding guide, although the second part of sentence is not precise—the words '*of those households*' refers to the first part of the sentence, *i.e*., to households that subscribe to cable TV. However, the number of households that own TVs is 12 times larger, not the number of households that subscribe to cable TV. Comparing this response with Query 5017 above, which was coded 11, hence allowing for a certain lack of precision, it was decided that the second part of the sentence was a mental leap rather than a logical error—code 11 was assigned.

Response P3 consists of two parts. The first part fits well the code 11 description and it is also similar to the sample Response 7 from the international training materials above (*The population of France is bigger than the population of Norway*.) However, there is also the second part, which does not fit the first part. Coders had to decide whether this is a case of contradictory elements, which would mean that the student did not understand the question. Or rather, is it a clumsy way to indicate that the percentages must be calculated with respect to the total populations in both countries? In the latter case it is just language clumsiness, caused by the lack of experience in formulating justifications. After a discussion, we decided to assign code 11—we gave the student the benefit of doubt.

Response P4 is one of the very few where the student was actually performing some calculations. But it is not just the calculation of the actual number of the cable TV subscribers, as in the code 12 description. It is rather an estimation used to justify a more general principle, formulated in the code 11 description. After a discussion we decided to assign code 12 to this response, to stress the presence of the calculations in the response.

In summary, difficulties in coding items that require an explanation or a justification of an opinion in most cases are caused by two factors. First quite often students' responses are much shorter and more laconic than those predicted by the coding guide. It is then hard to unambiguously conclude whether it fits the general code description. One can have doubts as to which of the following two cases takes place. Perhaps the student understood the claim and was able to justify it, but formulated the response in terms that were too general. Alternately, the student did not understand the claim or was unable to justify it and therefore offered a very general, ambiguous answer. On such occasions we need to draw a borderline between the level of generality that can be accepted as a correct answer and when it is insufficient. Second, students' responses are often ambiguous—they contain correct justifications and references to correct information, as well as parts that are not correct or simply hard to understand or interpret. Are those obscure fragments a result of the language clumsiness resulting from lack of experience in producing justifications? Are they caused by a mental leap or even by a language error? Or do they rather prove that the student did not understand the problem? On such occasions the coder's decision is rather subjective and depends on how a student's unclear response is interpreted.

A Second Example

Question 2 of PM00L Ice-cream Shop from the field trial for PISA 2012 is shown in Fig. 9.5 and its coding instructions are in Fig. 9.6. It belongs to a different type of item from the Cable TV question above, because it does not ask for an explanation but instead the student has to plan and perform calculations.

The "Show Your Work" Instruction

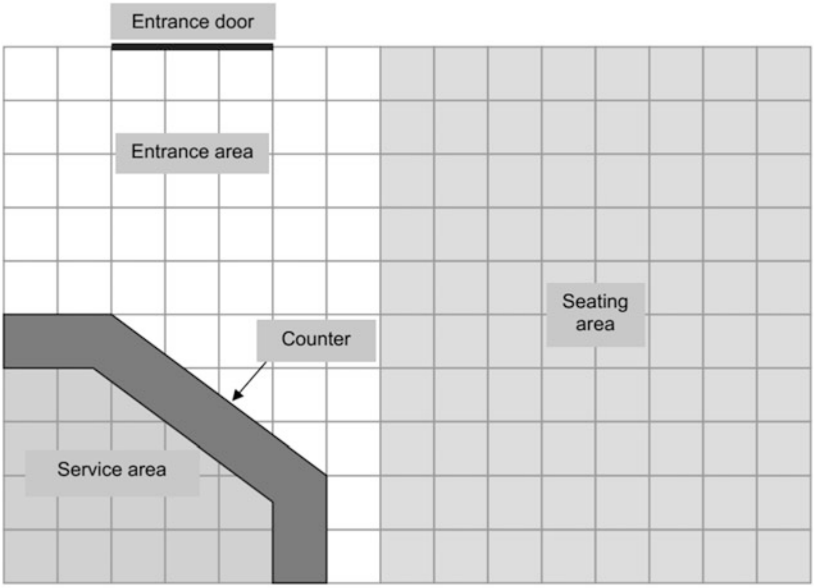
An apparent contradiction between the problem formulation and the coding guide is worth noticing. The item has an instruction "Show your work". However, from the code descriptions in Fig. 9.6 it can be seen that a student can obtain any code—including full credit (code 2)—even when he does not provide any calculations or show any working at all; it is enough to provide the correct answer.

ICE-CREAM SHOP

PM00L

This is the floor plan for Mari's Ice-cream Shop. She is renovating the shop.

The service area is surrounded by the serving counter.



Note: Each square on the grid represents 0.5 metres × 0.5 metres.

Question 2: ICE-CREAM SHOP PM00LQ02 – 0 1 2 9

Mari is also going to put new flooring in the shop. What is the total floor space area of the shop, excluding the service area and counter? Show your work.

(answer space omitted)

QUESTION INTENT:

Description: Calculate area for polygonal shapes
Mathematical content area: Space and shape
Context: Occupational
Process: Employ

Fig. 9.5 PM00LQ02 Ice-cream shop Question 2 with categorisation (OECD 2013)

Before we explain this rule of coding let us recall that the tested students solve problems in a dozen different booklets, assigned randomly to students. It is highly improbable that two students sitting next to each other would have the same booklets and solve the same item. Hence it is almost impossible that a student would copy the correct answer from another one, which would often be a danger in a classroom assessment. In these circumstances, we can safely assume that if a

Full Credit
Code 2: 31.5. [With or without units and with or without working. Note: It is likely that working will be shown on the grid. Incorrect units can be ignored because to get 31.5, the student has worked in metres.]
Partial Credit
Code 1: Working that clearly shows some correct use of the grid to calculate the area but with incorrect use of the scale or an arithmetical error. 126. [Response which indicates correct calculation of the area but did not use the scale to get the real value.] 7.5 x 5 (=37.5) – 3 x 2.5 (=7.5) – ½ x 2 x 1.5 (=1.5) = 28.5 m^2. [Subtracted instead of adding the triangular area when breaking total area down into sub areas.] 63. [Error using scale, divided by 2 rather than 4 to convert to metres.]
No Credit
Code 0: Other responses. Code 9: Missing.

Fig. 9.6 PM00LQ02 Ice-cream shop Question 2 coding guide (OECD 2013)

student has presented the correct answer then he is highly likely to have solved the problem unassisted. The student may have solved it mentally, or by performing a series of calculator operations or even by writing down the results of the intermediate calculations somewhere inside or outside the answer booklet, or erasing them.

What is then the rationale of including this instruction in the item? It is the following: if a student's answer is wrong but he follows this instruction and writes down the consecutive steps of the calculations, the coder gets the chance to track the steps and to find the reason for the student's error. Also, the coder can assign a partial credit to solutions containing computational error if such a code exists for the item. This would not be possible if only the answer had been provided, without any working.

Coding Difficulties with Items Requiring Calculations

The item PM00LQ02 Ice-cream shop Question 2 was one of the most difficult and laborious items to code in the whole history of PISA coding. It was obviously easy to assign code 2—to get it a student has to provide the correct answer of 31.5 without units or with units that are either correct or incorrect. It was much harder to decide whether the student's response deserves code 1 or 0. A necessary condition for receiving code 1 was the proper use of the square grid to calculate the area. Code 1 allowed for mistakes in scaling or calculation errors. To decide whether the student was properly using the grid, one had to monitor the reasoning path and this was very difficult. The number of ways that students chose to calculate this irregular area was practically infinite. Some divided the area into parts, most of them into two rectangles and a triangle or into three rectangles and a triangle. But quite often we encountered much finer dissections. Also different rectangles were used. Some started from a rectangle situated along the longer side of the ice-cream shop; others along the shorter side. The remaining part of the floor was divided into a large variety of different pieces. Some students did not add the floor area from simpler pieces; instead they subtracted from the total floor area the areas of the service area and the counter. But, of course, this could be achieved in many ways. Further variations arose because the student could calculate with the number of grid squares or measures in metres. Taking into account the possibility of making errors in counting the squares and/or computational errors, we obtain a huge number of possible combinations and hence of different solution paths. In this situation the attempt to determine whether a student correctly and consistently used the grid required genuine detective skills from the coders.

In some PISA mathematics items, calculations constitute only a part of the problem solution. Sometimes, besides performing calculations, the student has to interpret the obtained result. In other problems, before starting the calculations, the student has to find and understand the necessary data. Many different mistakes are possible when solving tasks of such complexity: improper or inaccurate reading of data, wrongly planned or performed calculations, wrongly interpreted results. Of course, any combination of the above is possible. Items of this complexity frequently have complex coding systems where different codes correspond to different categories of errors. In those cases, when the number of possible error combinations is large, coding is very difficult and requires of the coders a lot of effort, commitment and concentration.

At last, I would like to add a comment on the double digit coding. I believe that its potential still remains to be exploited. In the past administrations of Polish PISA we did not use this opportunity. However, the lesson has been learned: we have adopted the double digit coding in our education research on learning mathematics that lead us to very interesting conclusions concerning the way our students approach mathematics problems (Sułowska and Karpiński 2012).

Conclusion

There are many other interesting topics related to coders' work. During the last decade, while supervising the work of the Polish PISA mathematics coders' teams at the 2003, 2006, 2009 and 2012 assessments, my expertise was considerably strengthened. The coding process brings a lot of very detailed information about how students learn mathematics, which is much deeper than the codes reported to the international data base. I had the chance to use the experience gained in the consecutive assessments in the Polish core curriculum reform, as a Leading Expert for the Ministry of National Education, and later as an expert for the Polish Central Examination Commission helping to improve the quality of the national tests in mathematics. Recently, I have been involved in educational research at the Educational Research Institute, again capitalising on the lessons learned from PISA.

PISA is for me a fascinating adventure. I appreciate its guiding idea of mathematical literacy. PISA impresses me with its utmost diligence paid to the preparation of tools and procedures, including its great care for reliable coding. At all the stages of preparation, comments from the people involved in PISA around the world were appreciated. To my great satisfaction, also some of my comments about mathematics items and coding guides were considered useful.

References

Adams, R., & Wu, M. (Eds.). (2003). *PISA 2000 technical report*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2013). *PISA 2012 released mathematics items*. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf. Accessed 8 Oct 2013.

Sułowska, A., & Karpiński, M. (2012). Double-digit coding of examination math problems. *EDUKACJA an interdisciplinary approach*, 59–69. http://www.edukacja.ibe.edu.pl/images/pdf/eia-1-sulowska-karpinski-double-digit-coding.pdf. Accessed 23 Aug 2013.

Chapter 10
The Concept of Opportunity to Learn (OTL) in International Comparisons of Education

Leland S. Cogan and William H. Schmidt

Abstract Items addressing the Opportunity to Learn (OTL) construct, the idea that the time a student spends in learning something is related to what that student learns, was included in the mathematics portion of PISA 2012 for the first time. Several questions on the student survey were designed to measure students' opportunity to learn important concepts and skills associated with the assessed mathematical literacy. This chapter traces the development of this type of information in international comparisons of education and discusses four types of items that have been developed for this purpose. It also discusses the unique challenge of measuring this concept in PISA as it focuses on literacy, the knowledge students have acquired over their schooling to date, rather than on the content knowledge students have gained from schooling during a particular year or at a particular grade level. The specific OTL items and their purpose are identified from the Student Questionnaire section of Appendix A in the *PISA 2012 Assessment and Analytic Framework*.

An Opportunity Model

The mathematics portion of PISA 2012 included for the first time several questions designed to measure students' opportunity to learn important concepts and skills associated with the assessed mathematical literacy. The Opportunity to Learn (OTL) concept is the rather common sense notion that the time a student spends in learning something is related to what that student learns. This idea is fundamental to schools. As Bloom stated in his Thorndike award address, "All learning, whether done in school or elsewhere, requires time" (p. 682, Bloom 1974). Schools are created and organised to provide students with the time and selected learning experiences geared toward learning specific subject-matter content. The idea that the time spent learning something is important to what is learned is evident around

L.S. Cogan (✉) • W.H. Schmidt
Center for the Study of Curriculum, Michigan State University, East Lansing, MI 48824-1034, USA
e-mail: cogan@msu.edu; bschmidt@msu.edu

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_10

the turn of the last century in the writings of psychologists Edward Thorndike and William James (for a brief history of this, see Berliner 1990).

John B. Carroll, however, was among the first to feature time explicitly in his model of school learning (1963). Carroll posited that student learning was a function of both student factors: aptitude, ability, and perseverance; and classroom (or teacher) factors: time allocated for learning (OTL) and instructional quality. The latter is conceptualised as the interaction of the instruction provided with what is needed by the student in order to learn. Carroll summarised his model in the following equation:

$$\text{degree of learning} = f\left(\frac{\text{time actually spent on learning}}{\text{time needed to learn}}\right)$$

Carroll conceived of the ‘time actually spent on learning’ as the product of the ‘opportunity to learn’ provided by the classroom teacher and the student’s ‘perseverance’.

International Comparisons of Education

Comparisons of education systems by UNESCO, OECD, and others up until the early 1960s were primarily qualitative, consisting of rich descriptions of each national system. These descriptions often included tables of statistics that compared aspects of education that could be counted and quantified. Educational system characteristics and outcomes such as per pupil expenditures, teacher-pupil ratios, graduation rates, degrees, and proportion of students seeking further study were produced. However, what was yet missing was any measure of what students in each system might have learned or gained through their education experiences. The interest in exploring the creation of quantifiable measures that could be compared across systems was one of the impetuses that led to the creation of the Council of the International Project for the Evaluation of Educational Achievement (IEA) in the early 1960s. The council consisted of national education ministry representatives and university research professors who made plans to design and to conduct what came to be known as the First International Mathematics Study (FIMS) (Husén 1967; Travers and Westbury 1989).

Benjamin Bloom, who based his concept of mastery learning on Carroll’s model of school learning (Bloom 1974), was a member of the Standing Committee that was charged with leading and carrying out the project. Enough was known about differences in instructional practices and the curricula of the national systems represented in IEA to suggest that any measure of student learning or achievement was likely to vary substantially across the countries involved. Consequently there was also interest in measuring factors that might be related to such differences.

Stemming from Carroll's seminal model, which informed Bloom's mastery learning model, it was thought that

> one of the factors which may influence scores on the achievement examination was whether or not the students had an opportunity to study a particular topic or learn how to solve a particular type of problem presented by the test. (Husén 1967, pp. 162–163)

Although the 'opportunity to learn' (OTL) construct was conceived as operating at the individual student level, the challenges of a large-scale survey led to this being measured at the classroom (teacher) level through a teacher survey. In later analyses this simple index of students' OTL demonstrated a significant relationship with the achievement measures.

Although IEA studies have included descriptions of national education systems often including some of the same tables about the organisation of schools and schooling such as number of instructional days and teacher characteristics, the OTL index in FIMS demonstrated their focal interest on the teaching-learning process that occurs in schools. This was made explicit in the curriculum model introduced in the Second International Mathematics Study (SIMS) (Travers and Westbury 1989). This model articulated three instantiations of curriculum to be investigated: intended, implemented, and attained. The intended curriculum included the standards and expectations that education systems make known for student learning such as in curriculum frameworks. The implemented curriculum focused on classroom instructional practices and content. What students learned in school was represented in the model as the attained curriculum.

The IEA investigation of curriculum climaxed in the 1995 Third International Mathematics and Science Study (TIMSS). Prior to conducting the study early in 1995, a multi-year research and development project investigated the curriculum documents and classroom practices in multiple countries (Schmidt et al. 1996). This project produced curriculum frameworks for K-12 mathematics and science that were developed and adjudicated internationally. These frameworks were designed to be comprehensive of what any of the participating countries would teach in these subjects across the grades, and provided a common language for other aspects of the study thus yielding integrated curriculum measures. The frameworks were used to specify blueprints for the student assessments, classroom instruction topic categories in the teacher surveys, and the coding categories for the curriculum document analysis. National staff in each country trained by TIMSS document analysis staff coded their own curriculum documents. These included the official documents specifying what students were expected to learn at each grade (the intended curriculum) and a representative sample of textbooks used by students in the TIMSS targeted student populations. Textbooks embody a particular set of student learning expectations and provide resources to guide classroom instruction. Conceptually, textbooks form a bridge between what is officially intended for students to learn (intended curriculum) and the classroom instruction of teachers (implemented curriculum) becoming documents that give expression to a potentially implemented curriculum (Schmidt et al. 1996).

Countries used the different international benchmarks that TIMSS produced for each of these curriculum instantiations to inform various education reform efforts. Some were surprised by and dissatisfied with the large differences evident with what teachers reported teaching. Others were challenged by the curriculum expectations of other countries and used these to spur the development and formulation of new or revised curriculum standards. One example of the latter is the Common Core Standards for Mathematics recently adopted by a majority of U.S. states (Common Core State Standards Initiative 2010).

Literacy, Opportunity, and PISA

The prominent role of OTL in IEA studies is logical given their foundation in theories of student learning and the role that schools as organisations have in providing schooling (instruction) for students. The IEA curriculum model made explicit conceptual links between aspects of curriculum and the learning students attained through their schooling. This focus on school learning in IEA studies is evident in both the definition of the student population and in the sampling methodology. Student population definitions are grade focused as the question of interest relates to what students may know at a particular point in their schooling experience. Given the emphasis on student knowledge as a function of classroom instruction (schooling), these studies also gather information about classroom instruction from teachers. For these two curriculum indicators to be linked empirically the sampling of students and the sampling of teachers must be linked. Therefore, these studies sample entire classrooms in schools and survey the teachers of the sampled student classrooms.

The questions of interest in PISA have been less about what students know after studying a particular curriculum for a period of time, i.e., student outcomes at a particular grade level, and more on students' ability to use what they have learned through their accumulated schooling experience to address authentic, real-life challenges and problems. This practical orientation requiring the application of knowledge is the literacy that PISA has sought to assess. The difference in PISA focus and emphasis from IEA studies is expressed in both the definition of the student population and in the sampling methodology. The question PISA explores is what students of a particular age are able to do with the knowledge they have. This yields an age-based student population definition, i.e., 15-year olds, and a corresponding sampling methodology that is school based, randomly sampling students from a random sample of all the schools in which these students are to be found.

However, this shift in focus from the content knowledge students have gained from schooling during a particular year (grade level) to the application of the cumulative knowledge acquired over their schooling to date raises an interesting question: how relevant is opportunity to learn? How relevant is students' learning of core formal content-based competencies to their ability to apply their learning to authentic, real-world based problems and situations? Cognitive models of learning suggest that all learning is problem solving; the application and transfer of what has been learned in one context or situation to a different one (VanLehn 1989). Yet this does not clarify the specific types of OTL one might want to explore as being related to the literacy competencies measured in PISA. That school-based knowledge is related to PISA literacy seems clear from comparisons of results from TIMSS and PISA. For example, looking at the 26 countries/jurisdictions that participated in both the 2011 TIMSS and the 2009 PISA, the mathematics performance correlation was 0.87 (Mullis et al. 2012; OECD 2010). However, relative ranking on these two assessments did differ, sometimes rather dramatically, for some: a few did better on the TIMSS, e.g., the Russian Federation and Israel, and others did better on PISA, e.g., New Zealand and Norway. This similarity of results at least at the country level seems to suggest that the OTL issue, that is, the learning opportunities in schools, may well be pertinent for the development of literacy as assessed in PISA. What might literacy-pertinent OTL measures look like?

Traditionally, content or subject matter based OTL has been gathered through four different types of items. The first simply takes items from the student assessment and asks whether anything has been done in school that would enable students to obtain a correct answer on the test item. Response categories are typically binary, yes/no, but could also be expressed as some time gradient such as never, sometimes, and often. A variation of this method is to ask teachers to indicate how many students have had the opportunity to learn this type of problem. This was the method used in FIMS (p. 167, Husén 1967). A second option would be to present categories of school learning and to ask for a judgement of time each has been represented in schooling. Examples of mathematical experience could be formal school mathematics problems, mathematics word problems, problems involving the application of mathematics, and situations requiring the application of mathematics principles to real-world situations. This would simply yield an overall, relative indication of how much instruction time had been devoted to these various types of learning experiences.

A third option is abstracted from the first one listed above. In this option, exemplar problems that require the application of knowledge are presented and the respondent is asked whether anything like this has been done in school. This assesses more directly the extent to which students may have had experiences as part of their schooling in applying their knowledge in order to practise a particular skill. A final option is to present a full representation of subject-matter specific topics and ask to what extent these may have been encountered in school. With the PISA emphasis on assessing literacy, it seems likely that options one and three might be the most fruitful to explore

The PISA definition of mathematical literacy provides further guidance as to the specific aspects of students' OTL that are likely to be relevant. These are clarified in the PISA 2012 Mathematics Framework:

> 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. (OECD 2013a, p. 25)

The definition of mathematical literacy identifies specific skills to be assessed and, consequently, for which it would be appropriate to have some indication of students' OTL, i.e., some indication of what they may have encountered in their instructional experiences in school that would have helped them respond appropriately to the items or problems presented in the assessment. This identifies the information relevant to crafting the third type of OTL question described above. In addition, the Framework identifies four broad content areas for which some measure of student OTL would be appropriate: *Change and relationships*; *Space and shape*; *Quantity*; and *Uncertainty and data*. These four broad content categories provide an indication of the types of items that could profitably be used in an 'option one' type OTL measurement as well as defining the broad areas from which key topics/concepts might come for an 'option three' type OTL measurement.

PISA Measurements of Opportunity

In PISA 2012 the opportunity to learn measures were obtained through a series of items in the student questionnaire. The rationale for students providing their own OTL information is a function of PISA's age-based rather than grade-based methodology. PISA randomly samples 15-year old students from all classes in a school rather than sampling intact classrooms. Measuring OTL at the student level, however, is also consistent with the Carroll and Bloom models of student learning that first identified the OTL concept that in PISA and many other studies is considered an aspect of the learning environment. Most other comparative education studies have had teachers report on students' OTL. Although the PISA sampling methodology doesn't provide a way to estimate classroom effects, it does provide a true individual-level OTL measure that can be aggregated and analysed as a characteristic of schools and/or countries. Students' report on classroom instruction is sometimes criticised as unreliable as individual students in the same class tend to report differently and their reports do not always align well with what their teachers report. To the extent that interest in OTL is to explain student achievement rather than to reliably report on classroom instruction, the phenomenological response of the student may well be more powerful than a single teacher's report for multiple students. If a student can't recall encountering any sort of learning experience relevant to a particular topic or problem type this may indicate that the student does not have the needed knowledge to correctly solve the relevant item(s) or item type.

Six different items representing all but OTL 'option two' above were selected from the field trial and included in PISA 2012. This range of items across these three different approaches to the measurement of OTL represents a sort of

Q38 How often have you encountered the following types of mathematics tasks during your time at school? *(Please tick only one box in each row)*

		Frequently	*Sometimes*	*Rarely*	*Never*
a)	Using a <train timetable> to work out how long it would take to get from one place to another.	☐ 1	☐ 2	☐ 3	☐ 4
b)	Calculating how much cheaper a TV would be after a 30% discount.	☐ 1	☐ 2	☐ 3	☐ 4
c)	Calculating how many square metres of tiles you need to cover a floor.	☐ 1	☐ 2	☐ 3	☐ 4
d)	Understanding graphs presented in newspapers.	☐ 1	☐ 2	☐ 3	☐ 4
e)	Solving an equation like $3x+5=17$.	☐ 1	☐ 2	☐ 3	☐ 4

Fig. 10.1 Part of Question 38 (ST61) from PISA 2012 Student Questionnaire (OECD 2013a, p. 234)

'generalisability study' of mathematical literacy OTL. Each addresses the OTL issue for the application of mathematical literacy in different contexts. One item (question 38 in PISA 2012, see Fig. 10.1) asks students to indicate how often they have "encountered the following types of mathematics tasks" during their time at school. The nine tasks listed include a variety of formal mathematics tasks involving the application of mathematics knowledge in a real-world situation.

Another item (question 39) presents students with a list of 16 mathematics concepts (e.g. exponential function, divisor, vectors, rational number) and asks students how familiar they are with each one. The five response categories were: 'never heard of it', 'heard of it once or twice', 'heard of it a few times', 'heard of it often', and 'know it well, understand the concept'. Three of the listed concepts (proper number, subjunctive scaling, declarative fraction) were not true names of mathematics concepts to provide a check on a response bias (see p. 234, OECD 2013a).

A set of four items (question 44 through question 47 in Fig. 10.2) presented four different *types of problems* to students and asked them how often they had encountered such a problem type in: (1) their mathematics lessons, and (2) in the tests they had taken in school. The response categories for these were 'frequently', 'sometimes', 'rarely', and 'never.' Questions 44 and 47 each presented students with two examples of problem types requiring the application of mathematical skills or knowledge in a practical situation. Questions 45 and 46 each presented two

Q44 In the box is a series of problems. Each requires you to understand a problem written in text and perform the appropriate calculations. Usually the problem talks about practical situations, but the numbers and people and places mentioned are made up. All the information you need is given. Here are two examples:

1) Ann is two years older than Betty and Betty is four times as old as Sam. When Betty is 30, how old is Sam?

2) Mr Smith bought a television and a bed. The television cost $625 but he got a 10% discount. The bed cost $200. He paid $20 for delivery. How much money did Mr Smith spend?

Q45 Below are examples of another set of mathematical skills.
We want to know about your experience with these types of word problems at school. Do not solve them! *(Please tick only one box in each row)*

1) Solve $2x + 3 = 7$.
2) Find the volume of a box with sides 3m, 4m and 5m

Q46 In the next type of problem, you have to use mathematical knowledge and draw conclusions. There is no practical application provided. Here are two examples.
Example 1. Here you need to use geometrical theorems:

(image omitted – determine the height of a pyramid, given side lengths)

Example 2. Here you have to know what a prime number is:

If n is any number: can $(n+1)^2$ be a prime number?

Q47 In this type of problem, you have to apply suitable mathematical knowledge to find a useful answer to a problem that arises in everyday life or work. The data and information are about real situations. Here are two examples.

Example 1. PISA item Mathematics Unit 9 Robberies *(OECD 2009, p. 110)*
(item omitted)

Example 2. PISA item Mathematics Unit 46 Heartbeat *(OECD 2009, p. 148)*
(item omitted)

Fig. 10.2 Questions 44–47 (ST73—76) from PISA 2012 Student Questionnaire (OECD 2013a, pp. 235, 236) examples of problem types involving the use of formal mathematics content (see pp. 235–236 OECD 2013a). This sequence of four items measured how frequently students had the opportunity to work with word problems (Q44), applications of known rules and formulas (Q45), pure mathematics problems (Q46) and problems similar to previous PISA assessment items (Q47). The intention was to have students respond by considering the type of problem (as exemplified by the given mathematical tasks), rather than by considering the actual content such as solving equations or calculating percentages. Simple examples of each problem type were preferred for these items.

Summaries of results from the field trial for each of the items suggest that these various approaches to the measurement of mathematical literacy OTL will be of great interest in and of themselves. That is, the results vary across countries in a way that is of interest apart from any consideration of how this variation may be related to PISA mathematical literacy. In fact, in the initial PISA 2012 report one entire chapter documents the OTL variation across countries as well as some of the different ways OTL is related to PISA mathematics literacy performance (pp. 147ff, OECD 2013b). In addition, an OECD working paper reveals how the relationship between OTL and PISA mathematics literacy differs as a function of economic factors within countries (Schmidt et al. 2013).

Further analyses will no doubt reveal the fruitfulness of having included OTL as part of the PISA 2012 assessment. For example, such OTL items will enable researchers to explore issues of access and equity in educational opportunity within each country. As these have been included in a rather comprehensive survey, these issues may be explored additionally as a function of various measures of economic and social capital. It will also be possible to investigate the relationship between mathematics OTL and performance on the various PISA measures including the mathematics sub-scales and measures of reading and science performance. The particular way OTL is related to student-level and school-level socioeconomic measures is likely unique for each country as has been demonstrated with similar previously available data (Schmidt and McKnight 2012). The interpretation of these relationships and their attendant policy implications are also unique to an individual country's context.

References

Berliner, D. C. (1990). What's all the fuss about instructional time? In M. Ben-Perez & R. Bromme (Eds.), *The nature of time in school* (pp. 3–35). New York: Teachers College Press.

Bloom, B. S. (1974). Time and learning. *American Psychologist, 29*(9), 7.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*(8), 723–733.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, D.C. National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO).

Husén, T. (Ed.). (1967). *International study of achievement in mathematics, a comparison of twelve countries (Vol. II)*. New York: Wiley.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics report* (p. 516). Chestnut Hill: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Organisation for Economic Co-operation and Development (OECD). (2009). *Take the test. Sample questions from OECD's PISA assessments*. http://www.oecd.org/pisa/pisaproducts/Take%20the%20test%20e%20book.pdf. Accessed 14 May 2014.

Organisation for Economic Co-operation and Development (OECD). (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics and science* (Vol. 1, p. 276). Paris: OECD Publications.

Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy* (p. 264). Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD) (2013b). Measuring opportunities to learn mathematics. In *PISA 2012 results: What students know and can do—Student performance in mathematics, reading and science*. (Vol. I, pp. 145–174). Paris: OECD Publications.

Schmidt, W. H., & McKnight, C. C. (2012). *Inequality for all: The challenge of unequal opportunity in American schools*. New York: Teachers College Press.

Schmidt, W. H., Jorde, D., Cogan, L. S., Barrier, E., Gonzalo, I., Moser, U., et al. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Dordrecht: Kluwer.

Schmidt, W. H., Zoido, P., & Cogan, L. S. (2013). Schooling matters: Opportunity to learn in PISA 2012 (OECD education working papers, no. 95). http://dx.doi.org/10.1787/5k3v0hldmchl-en

Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula (Vol. 1)*. Oxford: Pergamon Press.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527–579). Cambridge: The MIT Press.

Part III
PISA's Impact Around the World: Inspiration and Adaptation

Introduction to Part III

This part demonstrates some of the ways in which PISA and its constituent ideas, methods and results have influenced education, drawing on the direct testimony of individuals many of whom have unique connections to PISA. The influence is of many types, including as a call to action from poor results, as a stimulus for new teaching and learning practices and for curriculum review, as a model for new assessment practices and provoking deeper education debates more generally and the creation of new educational standards. The underlying themes of the part are first of inspiration from PISA (both the need for change and possible directions for change), but second of adaptation of PISA resources, ideas and methods to meet the needs of very different educational environments.

This part is a collection of reflections on the impact that PISA has had on individuals, on education systems, and on teaching and learning practices in fourteen different countries. Inevitably, this is only a small sample of countries and a small sample of activities within the chosen countries. These reflections do not represent an official country view of the influence of PISA, and most importantly, they do not claim to represent all that is happening in these countries. Instead, they are written from the viewpoints of the authors and the initiatives with which they have been associated.

A striking feature of these reviews is the diversity of responses, including using PISA resources in teaching, to reviewing curriculum, through teacher education projects and formal assessment. This is a clear demonstration that aiming to improve educational systems requires working on many different fronts, and assessment results can stimulate many ideas for improvement.

Toshikazu Ikeda (Chap. 11) argues that the PISA Framework offers useful guidance to teachers on teaching mathematical modelling, on the selection or design of suitable problems, and significantly that particular modelling-related skills and competencies can be fostered through the kinds of problems used in

PISA. In support of this argument, Ikeda describes classroom practice that can advance relevant modelling skills. He also speaks from a broader perspective about why such practice is important, and concludes with a brief discussion of changes to Japanese curriculum resulting from local reflection on PISA results. Falling PISA results were one stimulus for the revision of the Japanese national curriculum to increase the time allocation for mathematics, and a two-pronged national assessment has been introduced, part focusing on basic skills and part on PISA-like problems.

Prenzel, Blum and Klieme (Chap. 12) give an overview of some of the significant impacts in Germany that have followed from their relatively poor mean performance in international surveys, especially the first PISA survey. A substantial impetus given to teacher professional development, the development and dissemination of new national performance standards for mathematics across several levels of schooling, and an intensified research focus on educational outcomes are key products of the PISA-related activity in Germany over the last decade or so. This contribution provides a clear example of concerted action leading to real improvements in educational outcomes over a relatively short timeframe, even within the constraints of a diverse federal system of government.

Arzarello, Garuti and Ricci (Chap. 13) provide a southern-European perspective on PISA's impact. As with the German example, below average national results in the early PISA survey outcomes have led to concerted action to improve educational outcomes particularly for the poorer-performing regions in the south of Italy. Beginning with information sharing, especially among teachers of PISA-aged students but extending to action at the precursor year levels, new approaches to curriculum and assessment have been introduced. They are supported by the development and dissemination of new classroom materials designed to foster the kinds of thinking valued through PISA.

Kai-Lin Yang and Fou-Lai Lin (Chap. 14) discuss some effects of PISA on educational practices in Taiwan, a perspective that differs from the previous two in that Taiwan has been a consistent high performer on international surveys such as PISA. They are concerned to improve from a high base. The article is focused on the selection of high achieving students by schools in their competitive and hierarchically structured system. Yang and Lin describe an attempt to use ideas underpinning PISA as the basis of a new selection system. The resulting debate has been studded with controversy regarding the relative merits of two goals that Yang and Lin refer to as 'learning power' (approached using a PISA-oriented curriculum and assessment) and mastery of textbook content characterised as 'mathematics for examination'. In Chaps. 13 and 14 and elsewhere in Part III, there is discussion about how the goals of a school curriculum (and hence the necessary assessment) are broader than PISA's mathematical literacy. Consistent with the goals of the PISA programme as set by the OECD, PISA mathematics derives its strength from a focus on the outcomes of the education that are most relevant to success in future life. However, mathematics as a school subject and as a branch of human endeavour is more than this. Consequently, these chapters discuss how the PISA framework needs to be broadened for a full assessment of school mathematics, particularly by including intra-mathematical argumentation and proof, ideas of mathematical structure, and mathematics motivated by interest and beauty, not only utility. The balance between mathematical literacy and intra-mathematical work in assessments will vary with the age and stage of students and, for those beyond the compulsory years, their purpose in studying mathematics.

Ten shorter pieces round out the discussion of the impact PISA has had in different countries, including nine countries that have participated in the PISA surveys, and one that has not. Almuna (Chile), Lindenskov (Denmark), Salles and Chesné (France), Zulkardi (Indonesia), Gooya and Rafiepour (Iran), Perl (Israel), Park (Korea), Kaur (Singapore), Rico, Lupiáñez and Caraballo (Spain), and Garfunkel (United States of America) provide a range of perspectives on important effects that PISA has had on educational debate and on classroom practice in their countries. Once again, these reflections do not claim to be comprehensive, and are not official reports. As well as the contributors being from variety of countries, there is great variety in their roles, from a teacher to the head of mathematics teaching for a country, people influential in teacher education, research, and curriculum development and people who have worked in the national agencies that contribute to PISA. There is also considerable variation in style of the accounts, ranging from quite official accounts to the intensely personal. In this chapter again, the themes of inspiration from PISA and diverse adaptation of PISA's ideas, resources and methods come to the fore.

Chapter 11
Applying PISA Ideas to Classroom Teaching of Mathematical Modelling

Toshikazu Ikeda

Abstract This chapter argues that the Mathematics Framework of PISA provides a meaningful guide for practical classroom teaching focused on mathematical modelling. The chapter discusses in detail how the Framework can provide guidance on choosing problem situations that interest students and also guide teaching students to appreciate the ways in which mathematics is used by society. In order to supplement the teaching of modelling through holistic problems involving all aspects of the modelling cycle, the chapter recommends the use of PISA-type problems to foster specific modelling competencies such as selecting variables and generating relationships. Advice on how this can be done is backed up by reports of experimental teaching. Finally, the effects of PISA in Japan are briefly discussed.

Introduction

Comparing test results among various countries in the world regarding mathematical literacy is one of the main purposes in PISA. However, PISA ideas can also make an important contribution to practical classroom teaching focused on mathematical modelling: firstly by considering the constructs and definitions that are set out in the Mathematics Framework, and secondly, by using sample PISA items as models for classroom tasks. This chapter discusses these two aspects. In particular, the definition of mathematical literacy, the four categories of contexts (*Personal*, *Occupational*, *Societal*, *Scientific*), and the three processes (*Formulat*e, *Employ*, *Interpret*) will be used as a guide when considering teaching plans aimed to foster students' competencies regarding mathematical modelling. This should be of value for teachers when setting teaching objectives, selecting a problem context, and introducing a problem situation. In the last section, there is a brief discussion on the treatment of PISA-type problems in Japanese classroom teaching. The suggestion is

T. Ikeda (✉)
Faculty of Education and Human Sciences, Yokohama National University,
Kinugasasakaetyo 2-68-42, Yokosuka 238-0031, Japan
e-mail: toshi@ynu.ac.jp

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_11

made that there needs to be more dissemination of ideas about how to encourage students to think deeply when they treat PISA-type problems.

Regarding teaching objectives, the definition of mathematical literacy from the PISA 2012 Mathematics Framework can be used as a guide to design a mathematics curriculum, a teaching plan, and so on.

> 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 judgements and decisions needed by constructive, engaged and reflective citizens. (OECD 2013, p. 25)

In the definition, two components can be seen. First is an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. This mathematical modelling capacity is a very important teaching objective in mathematics. Second is a citizen's recognition of the role that mathematics plays in the world and being able to use it in their lives. For students to recognise the role of mathematics in the world, it is necessary that the students have a lot of experience of solving real-world problems in a variety of contexts and additionally teachers must encourage students to reflect on the role of mathematics by comparing and contrasting those examples.

Regarding teaching methods, it is important for students to solve modelling problems completely so that they have experience in combining the different aspects that such problems require. They should have the opportunity to perform the whole modelling cycle, as illustrated in Chaps. 1 and 3 of this volume. On the other hand, it is also important for teachers to focus on the specific modelling competencies so that students can discuss and understand them and know how to use them. As everyone knows, it is hard for a teacher to treat the complete modelling process in the limited school time available. So one of the effective ways to use time is to set a problem that focuses on one or two of the constituent processes of mathematical modelling (*Formulate*, *Employ*, and *Interpret*) in the same way that many PISA items do. The case studies showed that the use of PISA items combined with group discussion and careful teacher direction was quite effective in helping to shape students' thinking about key features and stages of mathematical modelling. In this chapter, we will discuss how to apply the ideas of the PISA Framework in the classroom teaching of modelling.

Problem Situations to Interest Students

Problem situations that people are interested in differ according to the place where people are living, such as in which country and in what type of environment. It is obvious that the problem situations people face or are interested in differ between developing countries and developed countries. Further, even in the same country, familiar problem situations also differ between urban areas and suburbs. It is also said that problem situations that people are interested in differ between past society and present society. For example, constructing a figure to measure length or angle was important in the past but we now have convenient instruments to measure these things, so it is not as important now (Ikeda 2009).

In this respect, the four context categories (*Personal*, *Occupational*, *Societal*, *Scientific*) defined in the PISA 2012 Mathematics Framework (OECD 2013, p. 37) are useful to clarify when, where and for whom a problem situation is set. Before thinking about the teaching and learning of applications and modelling, it is suggested that the teacher understand the differences of problem situations so that he or she can plan that students will encounter a variety of situation to mathematise. In other words, it is not appropriate to focus on situations that only some of the students may have encountered outside of school. Drawing from the four different context categories will guide teachers to provide a balance by using a variety of problem situations.

When identifying the context, it is also important to consider place, time and person. As Jablonka noted, "different purposes may result in different mathematical models of the 'same' reality," and she gave an example about comparing mortgage plans.

> [For] the bank employee (aided by a software package), who must advise a client in the comparison of financing offers for a mortgage, for the manager of the bank this is a problem of profitability, and for the customer it is one of planning her personal finance. (Jablonka 2007, p. 193)

Further, the teacher should select an appropriate modelling task for teaching modelling. This suggestion raises practical questions, such as consideration of what is an appropriate modelling task. Galbraith (2007) makes two points regarding this question. First he notes the importance of consistency with avowed purpose. This is a basic and important issue that is sometimes neglected by teachers in practical teaching.

> If applications and modelling is included in mathematics education to attain goals such as 'students will experience school mathematics as useful for solving problems in real life outside the classroom' then students, to some extent, need to encounter tasks that are close parallels to comparable problem situations encountered outside the mathematics classroom. (Galbraith 2007, p. 182)

Galbraith (2007) also notes the importance of using models based on students' experience (which is influenced by their backgrounds) and the importance of motivation, which can come from "looking to the world and other disciplines for knowledge and problems" (p. 182). In considering these issues, there are different considerations for problem situations concerning the students at present, or in the future. If the problem situation concerns the present surroundings of students, is it concerned with most students or a few students? For example, the problem "What is the minimum size of a mirror where I can see my whole face?" and the problem of finding a strategy for "rock-paper-scissors" are in contexts familiar to most Japanese students, but problems about games such as soccer, tennis, and rugby are only familiar to students who are interested in these sports. If the problem situation concerns students in the future, is it concerned with them as citizens, as individuals or in their potential professional or vocational capacity? The former two situations concern many students. But occupational situations may only concern the particular students who want to work in that direction.

Fostering Specific Modelling Competencies with PISA-Type Problems

For the PISA 2012 survey, each of the questions was allocated to one of the following three processes, and performance on these was subsequently reported:

- *formulating* situations mathematically
- *employing* mathematical concepts, facts, procedures, and reasoning, and
- *interpreting*, applying and evaluating mathematical outcomes (OECD 2013, p. 28).

Problems involving these modelling processes can be seen in the assessment of modelling competency elsewhere. One example is Haines et al. (2001) and a very early example is Treilibs et al. (1980) who identified the five skills below that are especially involved in the formulating process and gave rich examples for teaching each of them:

- Generating variables—the ability to generate the variables or factors that might be pertinent to the problem situation
- Selecting variables—the ability to distinguish the relative importance of variables in the building of a good model
- Specifying questions—the ability to identify the specific questions crucial to the typically ill-defined realistic problem
- Generating relationships—the ability to identify relationships between the variables inherent in the problem situation
- Selecting relationships—the ability to distinguish the applicability of possible relationships to the problem situation (Treilibs et al. 1980, p. 29).

Treilibs's Sock Problem, shown in Fig. 11.1, is an example to test the skill of 'Selecting relationships', in this case in a graphical representation. The problem is to select a graph that shows a realistic relationship for socks that shrink in the wash. Students choose one graph from four. All of the graphs show the same total decrease in size (not numerically marked) and all four functions decrease monotonically. However the shapes of the graphs differ, so students have to think how the shrinking at each successive wash will relate to the amount of shrinking previously.

What will happen if we treat PISA-type problems that focus on distinct phases of modelling as a basis for classroom teaching about mathematical modelling? It was reported from a pilot study (Ikeda et al. 2007) that teaching using multiple-choice modelling problems focusing on distinct phases of modelling (e.g. Treilibs's Sock

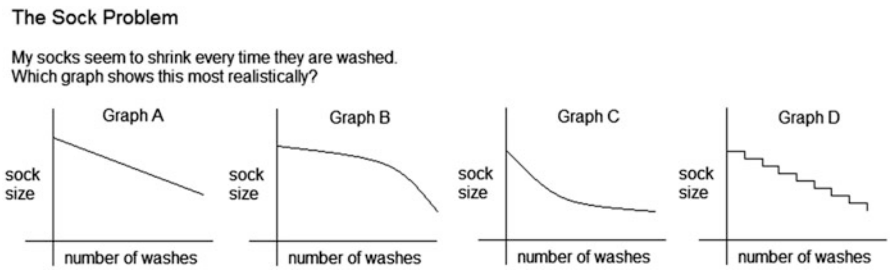


Fig. 11.1 The Sock Problem (After Treilibs et al. 1980, p. 35)

Problem above) can be a valuable teaching approach to foster students' thinking about modelling. These multiple-choice modelling tasks are, of course, no substitute for actually carrying out extended pieces of work involving mathematical modelling. But in many countries time to carry out such extended tasks is often hard to find in the crowded high school curriculum. Fully elaborated modelling tasks also present challenges for many teachers. On the other hand, multiple-choice tasks are familiar to teachers and students and may be useful in providing an introduction to mathematical modelling. These tasks should not be seen as ends in themselves. They can be used to provide students with an introduction to mathematical modelling, and can serve as a basis from which more serious work can proceed at a later stage. Here the teacher's role is crucial in keeping students focused on the larger picture.

A Teaching Experiment

Let us discuss the possibilities and limitations of using PISA-type problems in the teaching of mathematical modelling. This pilot study involved nine high school students in Japan divided into three groups of three members each. The empirical teaching was done in the following procedure. First, students solved the problems individually. Then they discussed their answers in their groups and after this answered three questions:

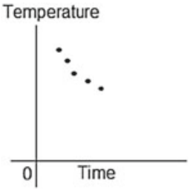
- Was your answer changed through discussion? Explain the reason.
- What kinds of issues were discussed?
- Justify your solution.

After this, all of the students discussed their answers together, with the teacher focusing the discussion on the most important issues. Then the teacher summarised the important ideas involved in solving modelling problems. For the study, all the students' discussions were recorded, and the transcripts were analysed.

We have chosen to consider the three multiple-choice modelling problems below and describe the students' performance. The Cooling Problem (see Fig. 11.2) and

Cooling Problem

On a warm summer day, some high school students decided to make a mathematical model to analyse how the temperature of a cup of coffee changes over time as it cools. By plotting the data of time and the temperature of the coffee, they obtained the graph below. Using the graph, the students investigated whether the following three types of functions could represent the relation between time x and temperature y of coffee (when $x > 0$).



Function 1 $y = ax + b$ (a, b constant)

Function 2 $y = ax^2 + bx + c$ (a, b, c constant)

Function 3 $y = ae^{-bx}$ (a, b constant)

Which **one** of the following explanations is **most** appropriate?

A. Function 1 is not appropriate because the temperature will become negative when the time goes by.
B. Even if we restrict the range of x, none of the functions are appropriate.
C. Even if we do not restrict the range of x, it is possible to use Function 2
D. Function 2 is not appropriate because according to this function the temperature of the coffee will eventually increase.
E. If Function 3 is used, the temperature is predicted to tend to zero as time goes by (provided the range of x is not restricted). However, it is possible to use Function 3 by transforming this formula.

Fig. 11.2 The Cooling Problem on generating and selecting relationships

the Mountain Problem (see Fig. 11.3) were developed by Ikeda et al. (2007, p. 103) and the Supermarket Problem (see Fig. 11.4) was developed by Treilibs et al. (1980, p. 31). The aim of the Cooling Problem is justification of a given model. The Mountain Problem focuses on selecting assumptions for the modelling to proceed. The aim of the Supermarket Problem is to focus on selecting variables.

The answers to the three problems given by each group are shown in Table 11.1. The students participated in the whole group discussion bringing to that discussion their own solutions and their reason for choosing those solutions. In the Supermarket Problem (Fig. 11.4), all the groups had chosen the correct answer. The teacher guided the students through a discussion of why this answer was correct. Even though some students initially gave an ambiguous reason that was not fully correct, after the whole group discussion, they clearly understood the reasoning behind the correct answer. In the Cooling and Mountain problems, only one group had a correct solution. For these two problems, at first the teacher guided students to distinguish shared opinions from individual opinions in the small groups. Then the

Mountain Problem
Consider this real world problem (do **not** try to solve it!).

It is impossible to see Mt. Fuji from Okinawa prefecture even if you have excellent eyesight. How far from Mt. Fuji is it possible for someone located at ground level to see it? Specify the distance from which it is possible to see Mt. Fuji using geometric arguments.

Which **one** of the following assumptions do you consider the **least** important in formulating a simple geometrical model to represent the problem situation?

A. Assuming the shape of Mt. Fuji as an equilateral trapezoid.
B. Assuming the shape of earth is a sphere.
C. Knowing sun's rays go in straight lines.
D. There is nothing to interrupt one's sight of Mt. Fuji.

To be able to see Mt. Fuji means to be able to see the upper half part of Mt. Fuji.

Fig. 11.3 The Mountain Problem focusing on making assumptions for modelling

Supermarket Problem
Consider this real world problem but do not try to solve it!

The management of a large supermarket is trying to estimate how many of its checkout tills should be operating at any given time. The factors or variables that could be taken into consideration include:

(a) average age of customers	(b) average bill size
(c) efficiency of the checkout girls	(d) maximum reasonable queuing time that can be expected of customers
(e) number of customers in the store	(f) average number of items bought
(g) pay rate for checkout girls	(h) proportion of customers using baskets rather than trolleys
(i) working hours	

Which **one** of the following sets of variables is **most** important in order to estimate how many of the checkout tills should be operating for customers?

A. (a), (c), (f), (i)	B. (c), (d), (e), (f)
C. (h), (c), (d), (e)	D. (e), (f), (g), (i)
E. (c), (e), (f), (h)	

Fig. 11.4 The Supermarket Problem focusing on selecting variables

Table 11.1 Answers to the three problems given by each small group

	Cooling	Mountain	Supermarket
Group 1	D	C	B[a]
Group 2	E[a]	D	B[a]
Group 3	A	A[a]	B[a]

[a]Correct answer

teacher guided students to first discuss shared opinions. Through active discussion, the students understood further issues and the different kind of ideas that had been put forward and sometimes derived new reasons that had not been discussed in their small groups.

Given below are transcripts of part of the whole group discussion of the Cooling and Mountain Problems, translated from Japanese. Through discussion, the students came to appreciate other issues and ideas, and learned how to evaluate other students' thinking. By exchanging ideas between groups, students made explicit the important ideas that are expected to be fostered in the teaching of modelling.

Partial Transcript of Cooling Problem Discussion

Teacher: Each group has a different answer for this problem; A, D and E. All groups rejected B and C. Why did you not select answers B and C?

Group 2: In answer B, if we restricted the range of x, it is possible to represent this phenomena with $y = ax + b$. Therefore, answer B is incorrect.

Teacher: How about answer C?

Group 1: Function $y = ax^2 + bx + c$ will increase when x is over a certain value. So, it is necessary to restrict the range of x. Answer C is incorrect.

Teacher: One of the groups has got the correct answer. (The teacher did not say which group.) Let's eliminate the other answers.

Group 3: We think that D is incorrect. Because when a in the expression $y = ax^2 + bx + c$ is a negative number, the value of y does not increase when x is getting larger. Therefore, the description "the temperature of the coffee will eventually increase" is not correct. Answer D is wrong.

Group 1: If you say "when a is a positive number," the value of y is never going down.

Group 2: This kind of argument is funny. For function $y = ax + b$, it is enough for us to restrict the value of a to a negative number. For function $y = ax^2 + \mathrm{b}x + c$, it is enough for us to restrict the value of a to be a positive number. It is enough for us to consider the case that fits the given situation.

Group 1: We have a question. How do you transform the function $\mathrm{y} = ae^{-bx}$ in answer E.

Group 2: The transformation means "+c". Namely the function becomes $y = ae^{-bx} + c$. If we set an adequate value for c, the temperature of coffee will converge to a certain temperature (that is room temperature) that fits the real situation.

Group 1: Can you show this by using a graphics calculator?

Teacher: Let me show you the graph of $y = ae^{-bx} + c$. (Presentation was made by the teacher using big screen at the front of the classroom.)

All: Great!

Teacher: How about A and D?

Group 2: In answer D, there is no description such as "restricting the range of x". If we set the range of x it becomes possible to represent the phenomena with $y = ax^2 + bx + c$. So D is incorrect. Further, if we restrict the range of x it is also possible to represent the phenomenon with $y = ax + b$. So answer A is also incorrect.

Teacher: Nice discussion! You have elicited some nice ideas. When we represent the phenomena by a function, we need to pay attention not only to the shape of the function but also the range of x. Further, we need to understand how the shape of function will change corresponding to changes to the coefficients of the function.

Partial Transcript of the Mountain Problem Discussion

Group 2: If the shape of earth was set as a plane, we can see Mt. Fuji from everywhere. As the shape is a sphere, there are areas from where we cannot see Mt. Fuji. So B is incorrect (i.e. it is an important factor to consider in the model).

Group 1 and 3: We agree with your idea.

Teacher: We can understand why B and E are incorrect. One of your groups has the correct answer. Let's eliminate the other groups' answers or provide a justification for the idea of your group.

Group 2: We think that C is incorrect. Because if the sun's rays curved, even though the shape of the earth was circular, we could see Mt. Fuji from everywhere.

Group 1: We see. We made a mistake.

Teacher: How about A and D?

Group 2: We thought D is correct. We cannot see Mt. Fuji if there is something in front of it. So it has no meaning to set the assumption that there is nothing to interrupt one's line of sight of Mt. Fuji.

Group 1: Group 2 is wonderful!

Teacher: What did you think, Group 3?

Group 3: We thought A is correct. There is no purpose to set the shape of Mt. Fuji as an equilateral trapezoid. It is possible to set the shape of Mt. Fuji as a triangle. Therefore, assumption A is meaningless.

Group 2: In assumption A, if a right triangle is placed on a circle, we can see Mt. Fuji differently corresponding to the placement of the right triangle. Therefore, the shape of Mt. Fuji is important.

Group 3: It is necessary to set the shape of Mt. Fuji. However, equilateral trapezoid is not important.

Group 3: I don't agree with the idea of Group 2. If we consider whether or not we can see Mt. Fuji at a certain place, it is important whether or not there is something to interrupt one's sight of Mt. Fuji. But, in this case, the problem asks how far from Mt. Fuji is it possible to see it. In other words, the problem is to find the length of the radius from the centre, at Mt. Fuji. Therefore, it is necessary to set the assumption that there is nothing to interrupt one's sight of Mt. Fuji.

Group 3: I have a thought about the previous idea of Group 2. If the shape of Mt. Fuji can be seen differently according to the direction, it is impossible to consider the geometrical problem in a two dimensional plane. If we consider the problem in a two dimensional plane, the shape of Mt. Fuji would be seen as the same from everywhere. Therefore, a right triangle is not appropriate in this case. Although it is necessary to set the shape of Mt. Fuji as a certain figure, it is not necessary to set the shape of Mt. Fuji as an equilateral trapezoid. For example, a triangle and a rectangle are also possible.

Group 1: At first, we agreed with the idea of Group 2. However, by listening to the idea of Group 3, I understand that the assumption "there is nothing to interrupt one's sight to see Mt. Fuji" is necessary. This problem asks about the possibility of seeing Mt Fuji, not whether someone can actually see it from a certain place.

Group 2: We understand.

Teacher: Nice discussion! As you discovered, the answer is A.

As shown in the transcripts of the two problems above, meaningful discussion took place between the groups, and students were able to elicit important ideas that promote modelling. On the other hand, we observed two limitations of the students' discussion that might be caused by using multiple-choice modelling problems. First, a few students tended to consider only how to eliminate the items, rather than to think about correct answers. This point will be shown in later analysis. As a result, students needed to be reminded that solving a real-world problem is not the same as checking and eliminating incorrect alternatives in multiple-choice answers.

Second, as multiple-choice modelling problems focus on the particular thinking that will be applied at a certain stage of the modelling process, it seemed that students tended to limit their considerations too strongly. For example, the Supermarket Problem given below is aimed at generating and selecting variables. In the partial transcript to be given below, the students only discussed whether each item was important or not. However, even in the stage of generating and selecting variables, we would like students to clarify the meaning of the given variables and also think about the relationships that might be generated between the variables. From the transcript that is given below, we can see that no students really clarified the meaning of the given variables when they were solving the multiple-choice problem.

In a classroom, an important teaching strategy is to treat the next or previous step of the modelling process after or during solving a multiple-choice modelling problem. In the transcript for the Supermarket Problem below, the teacher treated the next step of the modelling process, namely generating relationships immediately after selecting the variables. Students saw the importance of anticipating what kind of relationships might be generated when selecting variables. Tackling a whole modelling process by taking account of the different stages of the modelling process or analysing a certain stage of the modelling process by taking account of the whole modelling process are both important.

Partial Transcript of the Supermarket Problem

Teacher: All groups selected B. Why did you select B?

Group 1: We eliminated meaningless items. The average age of customers is irrelevant.

Teacher: Why do you think so?

Group 3: The aim of this problem is to estimate how many checkouts should be operated. So, if a customer was a child or grandfather, the age is irrelevant.

Group 1: Thank you for your assistance. The pay rate for checkout girls is also not important.

Group 3: Whether the checkout girl earned 1,000 yen per hour or 800 yen per hour, it has no bearing on the number of checkouts.

Group 1: Thanks again. The proportion of customers using baskets rather than trolleys is also not important. Some customers who buy small numbers of items choose to use a trolley. So this factor is not related to the number of items bought.

Teacher: Very good! This time, you could explain why the trolley choice is not important. Are there any more ideas?

Group 3: The average bill size is not important. Even if the bills of two customers were the same, the number of items bought could be quite different. It takes more time when the number of items bought is larger.

Group 2: The working hours is not related. After determining the number of checkouts, the working hours and number of checkout girls are determined.

Teacher: By eliminating the incorrect items (a), (b), (g), (h) and (i), each group was able to select the correct answer, namely answer B. Is there anyone who considered the relation between the four selected variables?

All: No.

Teacher: Let's consider the relation between the four variables.

Group 2: What is the meaning of the efficiency of the checkout girls?

Group 3: Let's consider that it means the time in seconds to check out one item. Let's ignore the time to put all the items into a bag. The unit is 'seconds per item'.
Group 1: How about the number of checkouts?
Group 2: I set the number of checkouts as x.
Group 1: How about the number of customers in the store? Some are selecting and taking goods and some are waiting for a checkout.
Group 3: We should set the meaning of this as the number of customers who are waiting for checkouts. The number of customers who are selecting goods is not relevant to the problem.
Teacher: Let's summarise the assumptions. (Teacher lists on blackboard.)

number of checkouts: x
efficiency of checkout girls (seconds per item): c
maximum reasonable queuing time: d
the number of customers who are waiting for checkouts: e
average number of items bought: f

Teacher: (Students work in groups and teacher resumes several minutes later). Let's explain the relation between the five variables.
Group 3: $ef/cx < d$
Group 2: $cef/d < x$
Group 1: Same as Group 2.
Teacher: Are these two answers the same or different? (i.e. from Group 2 and 3)
Group 1: Different. The location of c is different.
Teacher: Which is correct?
Group 1: We considered it by substituting concrete numbers in the formula. At first, the meaning of cf, namely multiplying 'efficiency of checkout girl (seconds per item)' by 'average number of items bought' is 'the time for one customer to pass through the checkout'. Next, the product of multiplying cf by e (the number of customers who are waiting for checkouts) gives 'the time for the last customer to wait for the checkout'. Then divide cef by d (maximum reasonable queuing time). As a result, each checkout till is assigned according to the maximum reasonable queuing time.
Teacher: Do you understand the meaning of dividing cef by d?
Group 3: No.
Group 2: I have another idea. Let's focus on the last person who is waiting for the checkout. As cef means the time that last person who is waiting to pass through the checkout, dividing cef by x that means the number of checkouts. Then we can get the time that the last person in each checkout should wait. This time should be shorter than d which is the maximum reasonable queuing time. By transforming the inequality $cef/x < d$ we can get the inequality $cef/d < x$.
Group 1: We considered a lot!
Teacher: What did you learn by formulating the inequality? What do you pay attention to when selecting variables?

Group 2: When selecting variables, we should check the meaning of variables, and anticipate the relation between variables.

Teacher: You made some very important points. Even though you can select important variables, this has no meaning if you don't formulate a relationship between them. It is important to clarify the meaning of variables by considering or at least imagining the relationship between the variables.

When combined with group discussion and careful teacher direction, the use of multiple-choice modelling tasks, as prepared by Haines et al. (2001), proved to be quite effective in helping to shape students' thinking about key features and stages of mathematical modelling in two relatively concentrated sessions. The problems in this pilot study were accessible and challenging to senior high school students of mathematics who had no prior teaching relating to mathematical modelling. Having a range of well designed and tested tasks on hand for teachers to use was a strategy that allowed students to come to terms with some important aspects of mathematical modelling within a relatively short period of time.

In this pilot case study, the teacher organised whole class discussion so that students could discuss shared ideas at first, then asked them to consider conflicting opinions from small groups by asking, "Why do you think the other group's idea is incorrect?" or "Why do you think your group's answer is correct?". In some cases, group discussion was able to bring all the students to a correct understanding of the problem. In other cases, by critiquing the ideas of their classmates and by listening to criticism, students realised that their explanation was still inadequate, ambiguous or unconvincing. As a result, they are pressed into giving clearer and more detailed explanations. The teacher's role was to help students identify the issues that need to be discussed, drawing on conflicting or opposing opinions among small groups, while not telling students the correct answer. When the teacher was unable at first to see opposing opinions among small groups, it was necessary for the teacher to probe students' thinking further so that conflicting or opposing ideas were exposed more clearly.

Purposes for Using Mathematics in Society

The definition of mathematical literacy (OECD 2013, p. 25) includes "recognising the role that mathematics plays in the world and making the well-founded judgements and decisions needed by constructive, engaged and reflective citizens" and states that the purpose of the mathematical thinking involved is to "describe, explain, and predict phenomena." These points are strongly concerned with the purposes for using mathematics in the real world. Niss (2008) has put the same ideas into slightly different words, when he identified three different kinds of purposes for using mathematics in other disciplines or areas of practice:

- In order to *understand* (represent, explain, predict) parts of the world
- In order to subject parts of the world to some kind of *action* (including making decisions, solving problems)
- In order to *design* aspects of the extra-mathematical world (creating or shaping artefacts, i.e. objects, systems, structures).

I think these three purposes help us to clarify the educational goals that students are expected to attain, the understanding of the modelling process for the beginner and the appreciation of the usefulness of mathematics in society. These three aspects are discussed in turn.

Educational Goals That Students are Expected to Acquire

This first point is characterised by the question: what kinds of educational goals are emphasised in teaching and learning mathematical modelling? Modelling is used for a variety of educational goals, such as foundations of science, critical citizenship, professional and vocational preparation, a way of living. There seems to be a strong connection between purposes for using mathematics and educational goals.

In the case of Niss's first purpose 'to understand parts of the world' and the 'predict, explain, describe' component of the PISA definition of mathematical literacy, parts of the world are considered to be phenomena of extra-mathematical domains such as nature or society. The mathematical model is verified by contrasting it with real data taken from the phenomenon being considered. Therefore, aims such as the foundation of science and professional or vocational preparation are emphasised more when we treat mathematical models that aim to 'understand'.

In the case of Niss's second purpose 'action' that references the well-founded judgements and decisions of the PISA mathematical literacy definition, parts of the world are considered problem situations, in which people have to make a decision or solve a problem. There are two types of mathematical model. First there is a social system model that is developed to make an objective and safe decision for people in a society, such as taxi prices or railway schedules. These models concern all citizens. After this mathematical model is embedded in a society, it becomes a main source for the reconstruction of reality (Skovsmose 1994). The second type is developed with personal purposes in mind, such as planning a family trip, or planning for family savings or loans. However, we must again note that "different purposes may result in different mathematical models of the same reality" (Jablonka 2007, p. 193). For example, trip planning may become part of a tour conductor's job. The mathematical model developed is effectively validated by developing another model to compare it with. Therefore, aims such as critical preparation for citizenship and for professions and vocations are emphasised more when we treat mathematical models that have the purpose of action.

In the case of ‘design’, which is Niss’s third purpose and again related to the well-founded judgements and decisions of the PISA definition, the focus is on objects that make our life more comfortable, such as furniture, architecture and designs using tessellation. This type of object is evaluated by an individual sense of value. Therefore, the aim of professional and vocational preparation is emphasised more when we treat a mathematical model that has a purpose to design. When we consider the teaching of modelling, we should examine the relation between the purpose for using mathematics and the educational goals that we have.

Understanding the Modelling Process for the Beginner

Considering Niss’s three purposes also helps us clarify the modelling process. The three purposes above imply that the modelling process depends on the purpose or the other disciplines. For example, when we understand a natural or social phenomenon, the mathematical model is abstracted from the real-world phenomenon, and also verified by contrasting it with real-world phenomena. However, when we make an action or design, multiple mathematical models are developed to make a decision, and the appropriate mathematical model is selected among several models according to the aim.

When we introduce mathematical modelling for students, a particular diagram (see examples in Chap. 3 of this volume) of the modelling process is often used to let students understand roughly what modelling is. We have to pay more attention to the fact that the modelling process differs according to the purpose for using the mathematics or the other disciplines involved, and teachers need to consider why they choose that particular modelling diagram with those students.

Appreciation of the Usefulness of Mathematics in Society

Third, Niss’s three kinds of purposes are also useful when we teach the usefulness of mathematics to students. When we teach how mathematics is used in a real-world situation, one of the methods is to identify purposes for using mathematics in the real world. By tackling a series of modelling tasks, students are expected to reflect on and find out the purposes for using mathematics in a variety of cases studied. For example, one of the methods is for the teacher to assess students’ appreciation of the usefulness of mathematics by asking “How is mathematics useful when we see real-world situations from a variety of viewpoints?” before and after modelling teaching. The teacher can assess how students deepened their appreciation of the usefulness of mathematics in a society, by comparing their writing before and after teaching modelling. For example, students’ writing can be assessed according to the viewpoint that the student takes. Writing at the first level is only from the students’ personal perspective. At the next level, it is from a social perspective, but it is not clear or only refers to special cases. At level 3, the social perspectives are clear and integrated, and may include the three different kinds of purposes identified by Niss and in the definition of mathematical literacy.

For example, the following responses (translated from Japanese) are from a Grade 9 student before and after experimental teaching, of 18 classroom periods of 50 min each (Ikeda 2002). Before the teaching, student A wrote:

> We can acquire mathematical thinking and judging from mathematics, but most people don't use mathematics in real life. So, it is not meaningful to consider how to use mathematics in real life in school.

This is assessed at level 2, because it adopts a social perspective. After the teaching, the writing is more elaborated and displays characteristics of level 3. Student A made progress regarding the appreciation of the usefulness of mathematics.

> Mathematics is useful to set the criteria or theory in a real world situation so that everyone can see what will happen. Mathematics is useful to consider before doing something. Using mathematics we can predict the result in advance without actually doing the thing.

Effect of PISA in Japan

In the PISA surveys of mathematical literacy, Japan was in the top position in 2000 (mean score 557), but its rank dropped to 4th in 2003 (mean score 528) and 6th in 2006 (mean score 523). This trend signalled the need for increased emphasis on mathematics and science in the recent revision of the Courses of Study. New Courses of Study for the elementary and lower secondary schools were announced in March 2008, and for upper secondary schools the change came in March 2009. The new Courses of Study were implemented in 2011 at the elementary school level, in 2012 at the lower secondary school level, and in 2013 in the upper secondary school level. In the new Courses of Study, time allocation for mathematics was increased.

In order to disseminate the spirit of the revised curriculum, national achievement tests and questionnaires were administered to all Grade 6 and Grade 9 students and their teachers from 2007 onwards. (A sample rather than the whole population was used in 2010–2012). There are two types of tests for students: one focussing on basic knowledge and skill, and the other targetting applications of mathematics. In the second type of test, the students are presented with problems similar to PISA tasks. These tasks test the ability to apply mathematical knowledge and skills in real-life situations and further test the ability to execute, evaluate, and modify a variety of plans to solve a given problem. The decision to disseminate problems like PISA tasks for all students at Grades 6 and 9 may be intended to change teachers' beliefs about the teaching of mathematics. On the questionnaires, elementary and junior high school teachers were asked how often they emphasised the relationship between mathematics and real-world situations. Possible responses included four

Table 11.2 Grade 6 teachers' responses to emphasis on real-world situations (percent)

	Often	Sometimes	Infrequently	Never
2007	8.8	51.4	38.4	1.4
2008	8.2	52.7	37.9	1.1
2009	7.9	54.1	37.0	1.0
2010	7.7	55.1	36.4	0.8
2012	7.4	56.0	35.5	1.0

Table 11.3 Grade 9 teachers' responses to emphasis on real-world situations (percent)

	Often	Sometimes	Infrequently	Never
2007	6.6	42.3	48.4	2.4
2008	6.0	43.8	47.3	2.7
2009	6.4	43.6	47.1	2.7
2010	5.9	44.9	46.3	2.8
2012	6.5	49.0	41.6	2.8

alternatives: often, sometimes, infrequently and never. The results of this question from 2007 to 2012 are shown as Tables 11.2 and 11.3 (National Institute for Educational Policy Research 2013). In 2011, this test was not implemented because of the great earthquake in the Tohoku area.

Tables 11.2 and 11.3 show that the relationship between mathematics and real-world situations is treated both in elementary and junior high schools. Teaching using PISA-type problems is also reported at both levels, although this kind of teaching is emphasised more in elementary than junior high school level. It is of concern that more than 40 % of junior high school teachers say that they seldom treat the relationship between mathematics and real-world situations. Anecdotal evidence suggests that some junior high school teachers believe that PISA-type problems do not make students think deeply, and that thinking deeply is better achieved by using intra-mathematical problems. This may be one of the reasons for the findings above. Consequently, there is a need for discussion and dissemination of ideas for encouraging students to think deeply when treating PISA-type problems.

Summary

It has been argued above that the Framework of PISA provides a meaningful guideline for practical classroom teaching focused on mathematical modelling. Three issues have been discussed in this article: choosing problem situations that people are interested in; fostering specific modelling competencies using PISA-type problems focusing on distinct phases of modelling; purposes for using mathematics in a society. Then the implementation of PISA-type problems in Japan has been briefly discussed. It is expected that because of the new Courses of Study and the new assessment, more extensive approaches using PISA-type problems and hence drawing on the PISA Framework may be implemented in classroom teaching in Japan.

References

Galbraith, P. L. (2007). Authenticity and goals—Overview. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI Study* (pp. 181–184). New York: Springer.

Haines, C., Crouch, R., & Davis, J. (2001). Understanding students' modelling skills. In J. F. Matos, C. Haines, R. Crouch, & J. Davis (Eds.), *Modelling and mathematics education* (pp. 366–380). Chichester: Horwood Publishing.

Ikeda, T. (2002). *The curriculum implementation focused on mathematical modelling in lower secondary school mathematics* (pp. 103–108). Paper presented at 35th conference on mathematical education, Japan Society of Mathematical Education (In Japanese).

Ikeda, T. (2009). Didactical reflections on the teaching of mathematical modelling—Suggestions from concepts of "time" and "place". In M. Blomhøj & S. Carreira (Eds.), *Mathematical applications and modelling in the teaching and learning of mathematics: Proceedings from topic study group 21 at the 11th international congress on mathematical education in Monterrey, Mexico, July 6–13, 2008* (pp. 217–228). Roskilde: Roskilde University Department of Science, Systems and Models.

Ikeda, T., Stephens, M., & Matsuzaki, A. (2007). A teaching experiment in mathematical modelling. In C. Haines, P. L. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics* (pp. 101–109). Chichester: Horwood Publishing.

Jablonka, E. (2007). The relevance of modelling and applications: Relevant to whom and for what purpose? In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 193–200). New York: Springer.

National Institute for Educational Policy Research. (2013). Results of National Achievement Test (In Japanese). http://www.nier.go.jp/kaihatsu/zenkokugakuryoku.html. Accessed 16 Aug 2013.

Niss, M. (2008). Perspectives on the balance between applications and modelling and 'pure' mathematics in the teaching and learning of mathematics. In M. Menghini, F. Furinghetti, L. Giacardi, & F. Arzarello (Eds.), *The first century of the international commission on mathematical instruction (1908–2008). Reflecting and shaping the world of mathematics education* (pp. 69–84). Roma: Enciclopedia Italiana.

Organisation for Economic Co-operation and Development (OECD). (2013). PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. Paris: OECD Publications. doi: 10.1787/9789264190511-en

Skovsmose, O. (1994). *Towards a philosophy of critical mathematics education*. Dordrecht: Kluwer Academic.

Treilibs, V., Burkhardt, H., & Low, B. (1980). *Formulation processes in mathematical modelling*. Nottingham: Shell Centre for Mathematical Education.

Chapter 12
The Impact of PISA on Mathematics Teaching and Learning in Germany

Manfred Prenzel, Werner Blum, and Eckhard Klieme

Abstract In this paper, various consequences of the PISA mathematics results in Germany are analysed. After a short review of the German PISA mathematics performance since 2000 the paper focuses on three aspects: fostering professional development of teachers, implementing educational standards, and providing empirical evidence by research programs. Altogether, PISA showed a strong impact on education in Germany and was an important stimulus for the discussion, reflection and improvement of the quality of mathematics teaching and learning in Germany as well as for research into mathematics teaching and learning.

Introduction

For decades Germany ignored international comparisons. The participation in such studies seemed to be superfluous, because almost everybody in this country was convinced of the high quality of mathematics and science teaching and learning in German schools. Nevertheless, a few educational researchers believed this accepted opinion should be challenged. They took the initiative for the participation of Germany in the Third International Mathematics and Science Study (TIMSS).

The decision to participate in TIMSS was worthwhile, because this assessment led to relevant insights. In mathematics the German students (Grade 8 secondary level)

M. Prenzel (✉)
TUM School of Education, Technische Universitaet Muenchen, Arcisstr. 21, 80333, Munchen, Germany
e-mail: manfred.prenzel@tum.de

W. Blum
Institute of Mathematics, University of Kassel, FB10 Mathematik and Naturwissenschaften, Heinrich-Plett-Str 40, 34132, Kassel, Germany
e-mail: blum@mathematik.uni-kassel.de

E. Klieme
Department for Educational Quality and Evaluation, German Institute for International Educational Research DIPF, Schlossstrasse 29, 60486, Frankfurt am Main, Germany
e-mail: klieme@dipf.de

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_12

attained 509 points on the TIMSS scale (Beaton et al. 1996; Baumert et al. 1997). Though this performance did not differ from the international average, Germany was really shocked—this was a widespread public perception.

This experience, however, was very salutary, with one immediate consequence. The educational policy authorities of all 16 federal states in Germany agreed to participate henceforth in international comparisons and in particular in the OECD Programme for International Student Assessment (PISA). They also decided to establish regular large scale assessments in Germany that should facilitate comparisons of educational outcomes between the federal states. In the first three administrations of PISA a systematic oversampling of schools and students allowed alignment of the educational outcomes of the German federal states on the PISA scale and comparison of performance both from a national and from an international perspective. Also, national expert groups extended the assessment frameworks and added national components to the test and questionnaire design. For example, the extended mathematics framework (Neubrand 2013) allowed for a deeper interpretation of the mathematics results in the various PISA cycles and also the identification of certain "profiles" within Germany. Thus, in Germany PISA became the renowned indicator for the quality of the school system and a synonym for top-quality assessment.

A Short History of PISA Mathematics Performance in Germany

The first PISA cycle (OECD 2001) taught Germany what 'shock' really means. At that time, Germany performed in mathematics (mean score M = 490, standard deviation SD = 103) significantly below the international OECD average (M = 500, SD = 100). The huge variation in student achievement, in particular the weak performance on the lower end of the distribution (5th and 10th percentile), large disadvantages for students with migrant backgrounds, and a very strong relationship between achievement and social background variables completed the impression of a real disaster. As all indicators showed severe problems, in newspapers some experts (especially from the OECD) predicted a dark future for Germany that could only be prevented by a complete reconstruction of the traditional school system.

Three years later, however, the picture looked somewhat different (OECD 2004) when Germany performed in mathematics at the OECD average (M = 503). Table 12.1 shows the further development of mathematics performance in Germany from PISA 2003 to PISA 2009 (Klieme et al. 2010). Finally in 2009 and 2012, Germany performed significantly above the OECD average. The students in Germany improved to an even greater extent in science (PISA 2000, M = 487; PISA 2009: M = 520). The increase in reading literacy has been moderate up to 2009 (PISA 2000: M = 484; PISA 2009: M = 497).

Table 12.1 Development of PISA mathematics performance in Germany from 2003 to 2012

	PISA 2003	PISA 2006	PISA 2009	PISA 2012
Germany average	503	504	513	514
OECD average	500	498	496	494

Many people attributed Germany's initial poor performance, among other factors, to its highly differentiated school system. From age 10, students are placed in different types of schools with different educational tracks. As basic structures of this differentiated school system have not been altered since PISA 2000, one can ask what other factors contributed to the improvement of mathematics performance (and science performance as well) in Germany during the last decade. Hence, are there lessons that can be learnt from the PISA history in Germany?

There is one general point that has to be kept in mind before going into details. The public reaction to PISA was absolutely sensational in Germany. PISA hit the headlines for weeks after the release of the results, especially after the first PISA administration in 2001, and this elicited enduring debates on the quality of schools in Germany. Thus, education moved much more into public attention. To date this interest and the awareness of the problem is still high. The detailed findings of the current fifth survey administration (PISA 2012) are awaited with great curiosity. It cannot be ruled out that this unique historical constellation is a general favourable condition for initiating and pursuing activities, measures and programs aiming at the improvement of educational processes and outcomes.

In the following sections, we will describe and discuss in more detail three reactions to, and consequences of, PISA that seem to have had an impact on the development of mathematics teaching and learning in Germany and hereby on the achievement progress attested over the course of the PISA survey administrations. These approaches represent exemplary efforts and measures at different levels of the education system aimed at an improvement of teaching and learning mathematics.

Fostering Professional Development of Teachers

The release of the TIMSS findings in 1997 first drew attention to possible weaknesses of mathematics teaching in Germany. The TIMSS Video Study (Stigler and Hiebert 1997) was especially helpful as it demonstrated vividly a monoculture of unimaginative mathematics tasks, activities and dialogues in German classrooms, with a focus on learning facts and procedures that are important for the next written test. The need for improvement of the prevalent style of mathematics teaching and learning was obvious, not only to experts from mathematics education, but also for many teachers, headmasters, supervisors and the authorities in the ministries. In Co-operation between the federal government and the federal states, a programme aiming at a prompt increase of the quality of mathematics and science teaching was launched (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 1997). The group of experts that was asked to develop a framework for this initiative decided to conceptualise a programme for the professional development of mathematics and science teachers. A thorough analysis of problem areas in the processes of mathematics teaching and learning and the underlying conditions (e.g., curricula, teacher training, teachers working is isolation) lead to the framework for this programme, which is referred to as SINUS (Prenzel et al. 2009).

At the core of the programme were 11 modules for improving teaching and learning, e.g. advancing the development of a "new culture" of mathematics tasks aiming at a much broader range of mathematical competencies (Niss 2003), securing basic understanding, and fostering cumulative learning in mathematics. Elaborated recommendations for teachers' activities contained in these modules helped the participating teachers to identify strengths and weaknesses of their teaching and provided examples and ideas for the development of advanced approaches. The programme intended to engage as many mathematics teachers as possible in SINUS, with teams in schools working continuously in a "module-oriented" way on the improvement of tasks, materials, and teaching approaches. Approved approaches from one team were first distributed and implemented within the school, and then distributed to other SINUS-schools in regional, and later national, networks of schools. The structure of modules helped the teachers to classify, interpret and integrate materials from other schools into their own teaching context. In the pilot-phase of the programme all these processes received various kinds of support from the scientific project staff and the scientific board of mathematics educators and researchers (e.g., examples of good practice, feedback, guidance, or special training). Step by step, SINUS produced a huge library of materials that from the beginning was made available for all interested teachers (via internet) or disseminated widely via manuals, books or teacher magazines.

SINUS started at the end of 1998 with 180 secondary schools and involved about 750 teachers. After a positive evaluation of the pilot phase (using also mathematics items and questionnaires from PISA) in 2003, the programme was expanded to 1,750 schools with a total of about 7,000 teachers. In 2004 a modified programme was offered for primary schools with a participation of 850 schools and about 4,500 teachers. Also after the end of the trial phase most of the schools continued the professional development in SINUS-Teams.

The SINUS programme was accompanied by a number of research projects (cf. Ostermeier et al. 2010). Besides an evaluation of the acceptance of the programme (which was high), research studies assessed how the teachers engaged in the programme and how they collaborated within schools and between schools. Different types of teachers with different needs for support were identified. Experts evaluated the materials and products that the teachers had developed. In the pilot phase the SINUS schools participated voluntarily in PISA 2000 and PISA 2003 (separately to the specified random sample for the official PISA survey). This design allowed comparisons with the national PISA sample and examined whether there was a selection effect operating in the recruitment of SINUS schools. The findings revealed no sampling bias: they were a typical sample of schools in

Germany. Students from SINUS schools described their mathematics lessons as much more cognitively challenging compared to students from the national sample. School and teacher questionnaires provided evidence for far more Co-operation among teachers in SINUS schools than in the national PISA sample. Student interest in mathematics as well as their self-concept was higher in SINUS schools. The mathematics performance tended to be higher especially for the weaker students. These findings provided evidence for the authorities to continue the programme for one decade and to scale it up in two phases of dissemination.

Was SINUS relevant to the German progress in PISA reported above? SINUS started at the end of 1998. In the proximate PISA 2000 assessment, the performance in mathematics and science was poor in Germany. It could not be expected that the fresh SINUS programme would have any impact on PISA 2000. But beginning with PISA 2003 the mathematics and science performance in Germany increased continuously. At the end, SINUS formally included 15 % of all secondary schools in Germany. Given the sampling procedures of PISA, however, it is unlikely that the increase in mathematics achievement can be ascribed only to the better performance of SINUS schools. Yet, SINUS did not only affect the schools involved in the programme. SINUS addressed relevant parts of the mathematics education community in Germany, and especially the group that is highly engaged in teacher training, in curriculum development, in publishing articles for teacher magazines or writing books. During the last decade materials for mathematics teaching and learning, like textbooks, sets of problems and exercises, recommendations and curricula, have changed considerably and most of them now reflect the SINUS modules and the joint philosophy of teaching and learning as well as of professional collaboration. So it can be assumed that SINUS did not only have an impact on the schools inside the programme, but also on schools outside the programme. It seems that both effects together could have been relevant indeed for the mathematics improvement in Germany over the course of PISA.

Implementing Educational Standards

The findings from PISA 2000 emphasised several additional challenges besides the average low performance in mathematics and the other domains. The proportion of very low performing students (on or below proficiency level 1) was nearly one quarter of the population of 15-year-olds in Germany. As PISA 2000 was already combined with a national oversampling to support comparison of the federal states of Germany, PISA revealed substantial differences between these states. In mathematics the gap between the best and the lowest performing states amounted to 64 points on the PISA scale (Baumert et al. 2002), which is equivalent to approximately two school years. All together, the PISA picture of Germany showed pronounced disparities in performance by region, social background, migration and gender. Also, by comparing PISA test scores to students' grades it was shown that grading standards varied considerably between states, and between schools within states.

The national supplement to the PISA 2000 mathematics test was based on an extended framework that aimed at completing the international PISA framework and at conceptualising mathematical achievement in a broader sense, for instance by taking into account also ‘technical’ aspects of procedural and factual knowledge, and by distinguishing between different “types of mathematical activities” (see Neubrand 2013, for details). By comparing tasks from state assessments in Germany to the international PISA test, it became clear that mathematics teaching in Germany had a strong focus on technical aspects of mathematics. By building a comprehensive model of mathematical competency, the national PISA 2000 report showed that complex modelling and problem solving—whether applied to everyday contexts or within mathematics—represent the highest level of mathematical proficiency (Klieme et al. 2001).

As all this was new information for the stakeholders, the lack of educational monitoring and of quality assurance in Germany became evident from benchmarking with successful PISA countries. A group of German researchers familiar with PISA was commissioned by the federal Ministry of Education to write a framework for the development and implementation of national educational standards in Germany.

This framework (Klieme et al. 2003) differentiated three related components of educational standards: educational goals, competency models, and corresponding assessment tasks. The suggestion was to conceptualise educational standards following this structure for all relevant subjects. The notion of competency models was very much based on the PISA experience. These standards were made obligatory for all federal states and all types of schools. Two different approaches of evaluation (regular formative evaluation at the school level, and regular assessment at the national level) were recommended to help provide feedback to teachers and monitoring information to the authorities.

As a prototype, national educational standards were developed for mathematics through a collaboration of mathematics educators and well-chosen teachers, conceptually based on the aforementioned mathematical competencies, which are also the conceptual basis of PISA mathematics (see OECD 2013 and Chaps. 1 and 2 in this volume). A second day of assessment linked to PISA 2006 was used to test the quality of standards-related tasks as well as for the scaling of items for a national mathematics assessment (Prenzel and Blum 2007). All these jobs were finished successfully. The obligatory national educational standards for mathematics were established in 2003 for the secondary level and in 2004 for the primary level. In the following years, standards-based recommendations for mathematics teachers were published (e.g., Blum et al. 2006) and a national centre for educational quality (Institut zur Qualitätsentwicklung im Bildungswesen—IQB) was established in Berlin providing tests for formative evaluation and organising national assessments based on the standards.

Altogether, these national educational standards certainly help teachers to get a clear focus on relevant educational goals and to understand the structure of (in our case mathematical) competencies and their cumulative development. The illustration of standards by tasks and the offer of tools for formative assessment support the teachers to identify strengths and weaknesses in the mathematical competencies of their students as well as the need for additional or different instructional approaches. With a longer-term perspective the regular comparative assessments at the national level are meant to contribute to a convergence of educational outcomes across the federal states in a positive sense. The intention is to continuously reduce the proportion of low performing students and the disparities in Germany and thus to raise the level of mathematical proficiency substantially and sustainably.

Providing Empirical Evidence

The findings of TIMSS did not only alarm stakeholders in Germany, but also groups of researchers in education. After the decision to participate regularly in future large scale assessments including PISA, different networks of researchers applied for the national project management. The commissioned PISA consortia in Germany included from the beginning distinguished researchers from all relevant fields and created networks of experts for the different domains (mathematics, science, reading). In particular for mathematics education as a research field, PISA had a special impact in Germany (see Bruder et al. 2013). The conceptual developments in this context contributed to a further development of mathematics education as a scientific discipline. These networks of experts were also engaged in other activities like SINUS and the development of standards.

Most important, however, was the development of a research agenda that used the different PISA survey administrations and samples as an opportunity to implement systematic research. The intention was manifold. The research ought to help to validate PISA and to provide additional evidence for the interpretation of the results of each PISA survey. Moreover, research projects ought to be linked to PISA to explore new methodological approaches. Finally, extensions of PISA with additional samples, target groups and follow-up assessments aimed at providing more solid evidence and at promoting basic educational research. Prenzel (2013) summarises some of these developments. It was very important to convince the authorities of the added value of these research programmes in order to get their approval and as far as possible also financial support. Assuring stakeholders of the need to support additional research was easier in the first phases of PISA when plenty of new and surprising information could be provided. In the beginning it was important to prove baseline information, such as checking that the international mathematics assessment is also fair from the perspective of German curricula and traditions. A second day of assessment allowed administering sets of items that represented different traditions and demands—and latent correlations with the PISA assessment above 0.90 were indeed found. At present, the expectations from the authorities tend more and more towards ideas and evidence for political decisions and actions that will both boost the performance of the students in Germany and simultaneously reduce disparities.

No matter how realistic these expectations will prove to be in the end, PISA is by its purpose and design insufficient for conceptualising educational reforms. To name but a few shortcomings, the cross-sectional PISA design constrains causal analyses; and age-based samples and questionnaires are insufficient to analyse the theoretically most important aspect, that is the quality of teaching processes in mathematics lessons. Thus, although the PISA tests and questionnaire scales are based on state of the art in research, the study design does not allow for sound conclusions on educational effectiveness (Klieme 2012).

With this background, several research initiatives were started in Germany to foster theory-driven educational research addressing more fundamental scientific questions. The issue of quality of educational processes and outcomes of schools was analysed from a systemic multi-level perspective in a priority programme funded by the German Research Foundation (Prenzel 2007). Quite a number of research projects in this priority programme were systematically linked to PISA (e.g., assessment of the impact of teachers' mathematics competencies, video studies of mathematics lessons, a longitudinal study of mathematics competencies before age 15). One example of these research projects was the so-called COACTIV Study (see Kunter et al. 2013) that proved, in particular, the relevance of different facets of the professional knowledge of mathematics teachers for quality instruction and for students' learning. Baumert et al. (2010) gives more details.

This 6-year priority programme on the educational quality of schools was followed by a new priority programme, also funded by the German Research Foundation, dealing with competence models (Hartig et al. 2008). The projects in this programme are analysing competence models for assessing individual learning outcomes in different domains both for students (e.g., student competencies in various mathematical subdomains, competencies in using pictorial representations, or cross-curricular problem solving) and for teachers (e.g., teachers' diagnostic competence). Projects are also analysing models that are suited for the longitudinal evaluation of educational processes. In the context of this program, Leutner et al. (2013) recently published an overview of concepts for modelling both summative and formative assessments with varying grain size. The priority program website provides more information http://kompetenzmodelle.dipf.de/en?set_language=en.

Concluding Remarks

In conclusion, PISA did have a strong impact on the public debate in Germany. Parallel to these public discussions manifold activities were started, and quite a number of these initiatives were carefully considered, well-orchestrated and substantiated by relevant research. PISA was an extremely important stimulus for the discussion, reflection and improvement of the quality of mathematics teaching and learning in Germany. Equally crucial was, however, the readiness of the authorities and the researchers to share their views and to start coordinated, evidence based programmes like SINUS or the development of educational standards. In gratitude for the recurrent stimuli from PISA a number of researchers from Germany try to bring ideas, suggestions and concrete work back to PISA.

References

Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., et al. (1997). *TIMSS—Mathematisch-Naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., et al. (Eds.). (2002). *PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske & Budrich.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill: Boston College.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (Eds.). (2006). *Bildungsstandards Mathematik konkret*. Berlin: Cornelsen.

Bruder, R., Barzel, B., Neubrand, M., Ruwisch, S., Schubring, G., Sill, H.-D., & Sträßer, R. (2013). On German research into the didactics of mathematics across the life span: National presentation at PME 37. In A. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 233–276). Kiel: PME.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (Eds.). (1997). *Gutachten zur Vorbereitung des Programms "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts"*. Materialien zur Bildungsplanung und zur Forschungsförderung, *Heft 60*. Bonn: BLK.

Hartig, J., Klieme, E., & Leutner, D. (Eds.). (2008). *Assessment of competencies in educational contexts*. Göttingen: Hogrefe & Huber.

Klieme, E. (2012). The role of large-scale assessments in research on educational effectiveness and school development. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 115–148). Dordrecht: Springer.

Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert et al. (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 139–190). Opladen: Leske & Budrich.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (2003). *The development of national educational standards: An expertise*. Bonn: Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research).

Klieme, E., Jude, N., Baumert, J., Prenzel, M., et al. (2010). PISA 2000–2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, & M. Prenzel (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 277–300). Münster: Waxmann.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (Mathematics teacher education No. 8). New York: Springer.

Leutner, D., Klieme, E., Fleischer, J., & Kuper, H. (Eds.), (2013). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Aktuelle Diskurse im DFG-Schwerpunktprogramm [Competence models for assessing individual learning outcomes and evaluating educational processes: Current discourses within the DFG priority program]. *Zeitschrift für Erziehungswissenschaft, Sonderheft, 18*.

Neubrand, M. (2013). PISA mathematics in Germany: Extending the conceptual framework to enable a more differentiated assessment. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on PISA: Research outcomes of the PISA research conference 2009* (pp. 39–49). Dordrecht: Springer.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papastavridis (Eds.), *3rd Mediterranean conference on mathematical education* (pp. 115–124). Athens: The Hellenic Mathematical Society.

Organisation for Economic Co-operation and Development (OECD). (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD.

Organisation for Economic Co-operation and Development (OECD). (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.

Organisation for Economic Co-operation and Development (OECD). (2013). PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. PISA, OECD Publishing. doi:10.1787/9789264190511-en

Ostermeier, C., Prenzel, M., & Duit, R. (2010). Improving science and mathematics instruction. The SINUS Project as an example for reform as teacher professional development. *International Journal of Science Education, 32*(3), 303–327.

Prenzel, M. (Ed.). (2007). *The educational quality of schools. Final report on the DFG priority programme*. Münster: Waxmann.

Prenzel, M. (2013). Research on PISA, with PISA, and for PISA. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on PISA. Research outcomes of the PISA research conference 2009* (pp. 1–12). Dordrecht: Springer.

Prenzel, M., & Blum, W. (Eds.). (2007). *Entwicklung eines Testverfahrens zur Überprüfung der Bildungsstandards in Mathematik für den Mittleren Schulabschluss. Technischer Bericht*. Kiel: Leibniz Institute for Science Education.

Prenzel, M., Stadler, M., Friedrich, A., Knickmeier, K., & Ostermeier, C. (2009). *Increasing the efficiency of mathematics and science instruction (SINUS)—a large scale teacher professional development programme in Germany*. Kiel: Leibniz-Institute for Science Education. Retrieved from https://www.ntnu.no/wiki/download/attachments/8324749/SINUS_en_fin.pdf. Accessed 23 Aug 2013.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Phi Delta Kappan, 79*(1), 14–21.

Chapter 13
The Impact of PISA Studies on the Italian National Assessment System

Ferdinando Arzarello, Rossella Garuti, and Roberto Ricci

Abstract In this chapter we sketch how the discussion that started in Italy with the disappointing results of the first PISA surveys was the origin of a national assessment program that possibly led to some improvement in the outcomes of mathematics learning. We will also underline similarities and differences between PISA studies and the Italian program of assessment.

Introduction

Discussion about the PISA program in Italy started, at least for teachers, from the 2003 results when Italy scored below the OECD mean. The teachers most engaged in innovative programs perceived the results as an alarm bell concerning the state of teaching and learning in Italian schools at the end of the compulsory cycle of schooling, which in Italy ends at age 16 years. It is interesting to consider the changes (if any and of what nature) in the PISA results for mathematics in the subsequent years. In fact, some elements have not changed. The Italian mean scores (466 in 2003, 483 in 2009) continue to be below the OECD mean and there is a great variability between the Italian regions. Specifically, while in northern regions there are results above the OECD mean, the opposite happens in the southern regions. However, as shown in Fig. 13.1, from 2003 to 2009 in mathematics there was a positive trend, with an increase of 17 points (0.17 standard deviations). Figure 13.2 shows the mean scores for five areas, using data assembled from PISA reports. It reveals that this better performance is due above all to the better results in the southern regions, particularly from 2006 to 2009. Regions in the Sud area improved by 25 points and regions in the Sud Isole area improved by 34 points. Even though they remain below the OECD mean, they show better performance. Let us try to explain this change.

F. Arzarello (✉)
Department of Mathematics, University of Turin, Turin, Italy
e-mail: ferdinando.arzarello@unito.it

R. Garuti • R. Ricci
Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione - INVALSI, Via Francesco Borromini, n. 5 – 00044 FRASCATI, Rome
e-mail: rossella.garuti@libero.it; roberto.ricci@invalsi.it

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_13

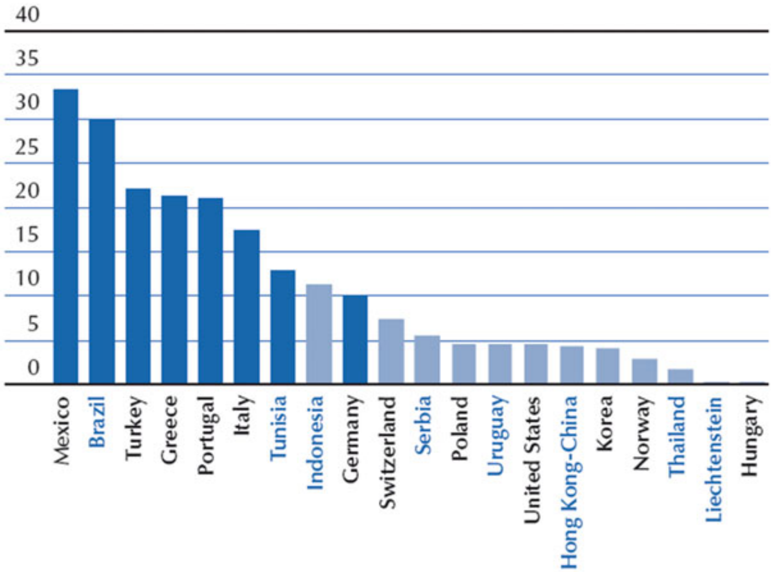


Fig. 13.1 Score point change in mean mathematics performance between PISA 2003 and PISA 2009 showing those countries that improved (Adapted from Fig. V.3.1, OECD 2010, p. 60)

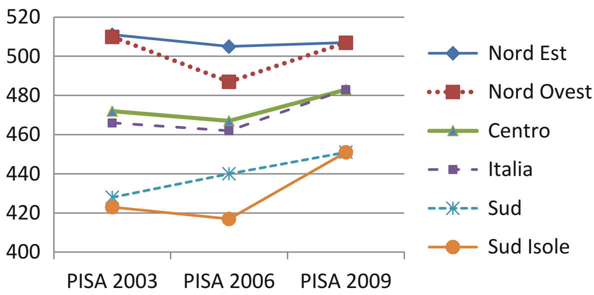


Fig. 13.2 PISA Mathematics mean scores for the five areas of Italy

In-Service Teacher Education About PISA

The Plan for Information and Awareness

Because of the 2003 and 2006 PISA results, the Italian Ministry of Education (MIUR) in 2008 launched the program ‘Piano di informazione e sensibilizzazione sull’indagine OCSE-PISA e altre ricerche internazionali’ (‘Plan for information and awareness about the OECD-PISA study and other international research'). The program has been funded with European money and its aim is supporting innovation and quality of teaching in the schools of four southern Italian regions (Calabria, Campania, Puglia and Sicilia) in order to bridge the gap measured by PISA with respect to both other Italian regions and to the states of the European Union. They were chosen because they have a per-capita GDP less than three quarters of the mean for the European Union. These regions are in the areas Sud and Sud Isole of Fig. 13.2.

The program started in 2008–2009 and involved the teachers of Italian, mathematics and science in the first 2 years of upper secondary school (Grades 9 and 10) in all schools in the four regions (altogether 20,000 teachers). The program consists of a 2-day seminar, and provides materials that the teachers have to study and discuss together when back in their schools. The main goals of the program are:

- Informing teachers about the OECD-PISA study in a clear and correct way
- Analysing the PISA Framework for mathematics, particularly the structure of the test and the public items
- Comparing them with the most common didactical practices in Italian classrooms
- Analysing the results of Italian students in the PISA study.

As the project progressed, it became apparent that the mathematical competences considered by PISA are not the exclusive concern of the Grade 9 and 10 teachers, whose classes are directly involved in PISA testing, but they must be built up over longer periods of time, starting from the very beginning of school. Hence, after 2009 the project was enlarged to the teachers of primary and lower secondary schools: currently it is targeting teachers of Grades 6 to 8.

The m@t.abel Project

The changes in Italian teaching practices that have been stimulated through PISA may have been caused also by another project. The m@t.abel project is a big teacher education program, promoted by the MIUR from 2008. This is an acronym that in Italian means *basic mathematics with e-learning*. Teachers were divided into virtual classes of 20 persons under the guidance of an experienced trainer, to share the materials of the course (about 80 examples of teaching activities in the classrooms) and discuss what happens in their classrooms when they trial the teaching units of the project. It has involved more than 5,000 teachers from Grades 6 to 10 all over Italy, included some of the teachers from the same four southern Italian regions listed above. The main aim of m@t.abel consists in providing examples of best practices in the classroom, and these are often well aligned with the PISA Framework. We do not have the space to discuss it here. The reader can find more information from the project booklet (Arzarello et al. 2012), which makes explicit the relationships between the Italian project and the PISA study (pp. 22–25).

Many of the teachers of the other program also participated in this one. They found in the m@t.abel materials many concrete examples of activities that enacted in the classroom what is stated theoretically in the PISA Framework. The two programs have involved almost all teachers in southern Italy, where the change in PISA results has been more dramatic.

The Italian Assessment System

Over time, an additional topic has been added to those covered by the seminar: the PISA Framework is now compared with that used by the Italian Assessment System (SNV: Servizio Nazionale di Valutazione). SNV started its work in 2008 through annual surveys conducted by the National Evaluation Institute for the School System (INVALSI) at different school grades. The home page for INVALSI is http://www.invalsi.it/invalsi/index.php. The INVALSI develops standardised national tests to assess pupils' reading comprehension, grammatical knowledge and mathematics competency, and administers them to the whole population of primary school students (Grades 2 and 5), lower secondary school students (Grades 6 and 8), and upper secondary school students (Grade 10).

As well as participating in PISA, since 1995 Italy has also participated in the TIMSS program, which measures students' competencies in mathematics and science at Grades 4 and 8. The results of the Italian students have been very disappointing: especially in 2007 when the ranking of Italy in TIMSS decreased dramatically. But something new for Italy happened from 2008: from that year all students in Grade 8 had to face a national final standardised SNV test on reading and mathematical competence at the end of lower secondary school in addition to the normal final examination organised by the school. Up to that date, nothing similar existed. In the next TIMSS testing conducted in 2011, Italy was the country with the greatest improvement in mean score from 2007 to 2011. With this improvement, Italy reached the international TIMSS mean.

Of course it is too crude to postulate a cause-effect link between the introduction of the Italian Assessment System and this tangible improvement. However such a conjunction is a fact and this event is the only real change that happened in Italian schools in the period 2007–2012. It is more than an impression that the introduction of the standardised tests at the end of the lower secondary school has represented a strong innovative component, which has produced innovation and a revision of the practices in the schools. Certainly more investigation is needed to understand these results but there is no doubt that the introduction of standardised tests has been a strong element to trigger and support the revision of the teaching methods adopted by teachers in schools. Furthermore the results of PISA 2012 will represent for us a particular element of interest in order to understand the fallout of the activities implemented in the schools described above.

Frameworks for the Italian Assessment System and PISA

The PISA Mathematics Frameworks (OECD 2004, 2013) have certainly influenced the construction of the Reference Framework for Mathematics of the Italian Assessment System, SNV. Its investigation aims to take a snapshot of schooling as a whole: in other words, it is an evaluation of the effectiveness of education provided by Italian schools. Currently, standardised tests are administered every year to all students at five grade levels from Grade 2 to 10, and within the next 2 years, to Grade 13 as well. As noted above, the Grade 8 test is included in the final examination at the end of the first cycle of instruction: its main aim is providing teachers with a nationally benchmarked tool for the assessment of their students. The results of a national sample are annually reported, stratified by regions and disaggregated by gender, citizenship and regularity of schooling. These results are public, as well as the tests and the marking schemes. However, the results of each school are sent confidentially to the principal. From 2013, some items are kept secure and used to anchor the results over time. There are at least three main differences between SNV tests and PISA surveys: the frequency (annual vs. triennial), the type of tested population (census vs. sample) and the chosen population (grade-based vs. age-based students).

The preparation of the SNV items is performed in two steps. A first set of items is prepared by in-service teachers of all levels, who also classify them according to the SNV framework (question intent, processes involved, precise links with the National Guidelines). Subsequently, the SNV National Working Group builds the test by selecting items so that the test is balanced both from the point of view of content and of processes. However, the methodological and statistical methods underpinning SNV and PISA are basically the same. The Reference Framework for Mathematics in SNV has its roots in the National Guidelines for the Curriculum and in some teaching practices that have consolidated over the years. Another important reference is the UMI-CIIM curriculum "Mathematics for the citizen" (Anichini et al. 2004), which is based on results of mathematics education research and has deeply influenced both the last formulation of the national curriculum and the m@t.abel program. "Mathematics for the citizen" explicitly states the necessity of taking into account

> both the instrumental and the cultural function of mathematics. [. . .] Both aspects are essential for a balanced education. Without its instrumental features, mathematics would be pure manipulation of signs without meaning; without a global vision mathematics would be a series of recipes without method and justification. (Anichini et al. 2004, p. 7, translated by authors)

The SNV Framework defines what type of mathematics is assessed with the SNV tests and how it is evaluated. It identifies two dimensions along which the questions are built:

- The mathematical content, divided into four major areas of Numbers, Space and Figures, Relations and Functions, Data and Forecasts
- The processes that students should activate while solving the questions.

This subdivision of content into four main areas is now shared at the international level: in PISA there are four content categories (*Quantity*, *Space and shape*, *Change and relationships*, *Uncertainty and data*) and in TIMSS there are four content domains (Number, Geometry, Algebra, Data and chance). As one can see, the differences are minimal and the four areas broadly identify the same categories of mathematical content, even if one can observe different choices according to what kind of mathematics the items are assessing. The Italian choice has been to name areas by the mathematical objects involved and not by the academic name of the discipline, which has its own well defined epistemological status (e.g. Space and Figures and not Geometry). This choice by SNV matches the National Curriculum but is a departure from tradition.

Concerning the processes, we note that the PISA 2012 Framework (OECD 2013) more so than the PISA 2003 Framework (OECD 2004) moves towards this direction with a definition of mathematical literacy focused on the mathematisation/modelling cycle (see Chap. 1 of this volume). In order to choose items and to analyse results, the SNV study considers the following types of capabilities:

- Knowing and mastering the specific content of mathematics
- Knowing and using algorithms and procedures
- Knowing different forms of representations and passing between them
- Solving problems using strategies within different areas (numerical, geometrical, algebraic, etc.)
- Acknowledging the measurability of objects and phenomena in different contexts, using measuring tools, measuring quantities, estimating such measures
- Using typical forms of mathematical reasoning (conjecturing, arguing, verifying, defining, generalising, proving,. . .)
- Using tools, models and representations in the quantitative treatment of information from scientific, technological, economic and social environments
- Recognising shapes in space and using them to solve geometric or modelling problems.

Starting from 2013, SNV adopted a further classification, namely the same used by PISA (*Formulate—Employ—Interpret*) in order to allow an easier comparison of the two surveys. The definition of mathematical literacy in the PISA 2012 Framework (OECD 2013) is centred more on the idea of mathematics as a means to analyse, interpret and represent real-word situations (the cycle of mathematisation/modelling). However the framework adopted by SNV assessment is strictly connected to the *national curriculum* and includes aspects of mathematical modelling as in PISA, and aspects of mathematics as a body of knowledge logically consistent and systematically structured, characterised by a strong cultural unity (Anichini et al. 2004). The two examples below highlight these aspects. The SNV Mathematics Framework is a tool in evolution, in the sense that periodic updates are to be expected, based on experiences from the testing and input from schools.

Two Examples from the SNV Study: Mathematical Modelling and Argumentation

We sketch here two examples in order to highlight similarities and differences between the SNV and PISA frameworks and in the way mathematics is considered in the two studies.

The Elongation of a Spring

The first example (see Fig. 13.3) is a question involving mathematical modelling used in SNV 2011. Two versions with the same stem but differing in the multiple-choice options offered were used: one for Grade 8 and one for Grade 10. Both items are within the area *Relations and functions* and concern mainly the capability of *using tools, models and representations in the quantitative treatment of information* (the seventh capability in the list above). To answer correctly, students must interpret the meaning of the parameters of the function (L_0 and K) in terms of the physical characteristics of short and hard.

Table 13.1 shows the overall results from the national report (INVALSI 2011). It is not very surprising that more Grade 8 students than Grade 10 students are correct, for at least two reasons. First the values of the parameters are different and those for Grade 8 are easier to compare. Second, it is usual in lower secondary school to represent physical phenomena through formulas and graphs, while this is generally done in secondary school only after Grade 10. In Grades 9 and 10 algebra is

The formula $L = L_0 + K \times P$ expresses the length L of a spring as P, the weight applied, changes. L_0 represents the length in cm of the spring at rest and K indicates how much the spring stretches in cm, when a unit weight is applied to it.

Which of the formulas below represents better the following description?
(Hard means that the spring is very resistant to pulling)

D17 (Grade 8 version)	D24 (Grade 10 version)
It is a very short and hard spring.	*It is a very long and hard spring.*
A. $L = 10 + 0{,}5 \times P$	A. $L = 15 + 0{,}5 \cdot P$
B. $L = 10 + 7 \times P$	B. $L = 75 + 7 \cdot P$
C. $L = 80 + 0{,}5 \times P$	C. $L = 70 + 0{,}01 \cdot P$
D. $L = 80 + 7 \times P$	D. $L = 60 + 6 \cdot P$

Fig. 13.3 Relations and functions items from SNV (2010–2011)

Table 13.1 Percentage of students choosing each option in the national sample (INVALSI 2011)

Item	Options				Omissions
	A	B	C	D	
D17 (Grade 8)	58.3[a]	25.4	7.9	4.3	4.0
D24 (Grade 10)	8.1	33.2	38.1[a]	8.9	11.8

[a]Correct answer

E13. The teacher asks: "An even number greater than 2 can always be written as the sum of two different odd numbers?"
Below are the answers of four students.
Who has given the correct answer? Justify it properly.

Antonio:	Yes, because the sum of two odd numbers is an even number.	(44.0%)
Barbara:	No, because 6 = 4 + 2.	(6.4%)
Carlo:	Yes, because I can write it as the odd number that precedes it, plus 1.	(34%)*
Daniela:	No, because every even number can be written as a sum of two equal numbers.	(14.0%)

*correct

Fig. 13.4 Item E13 from SNV (2011–2012) at Grade 8 (with percent choosing each option)

generally taught only at the syntactic level, at most to solve geometric problems and never to model physical situations, which is left to Grades 11, 12, 13 (Garuti and Boero 1994).

Natural Numbers: Justifying and Proving

The example in Fig. 13.4 arises in the context of the latest Italian research in mathematics education (Mariotti 2006; Boero et al. 2007). It somehow condenses the results of wide research about the approach to argumentation and proof in mathematics, even with young students. Such research has important implications in the field of educational research, and also suggests strongly innovative teaching practices in the classroom. The example is classified in the area Numbers and relates to the sixth capability *(using typical forms of mathematical reasoning)* in the list above. In this item, Grade 8 students are required to evaluate arguments about the validity or non-validity of a non-trivial statement: they must choose the answer that shows the correct justification. This item requires that the student understands that every even number can be written as $(2n-1)+1$. In case of the number 2 the formula still holds, but the sum is between two equal odd numbers.

The chosen distracters correspond to the more frequently observed behaviours of students in the research quoted above: they all concern students' understanding and exploration of the statement. In particular, the distracter A, which had 44 % of responses, corresponds to an inversion between the thesis and hypothesis: to answer the question it is not relevant that the sum of two odd numbers is always even. We consider questions of this type very important since:

- Within a standardised test, they assess verifying mathematical skills that are typical of the cultural aspect of mathematics;
- They show teachers the possibility of using algebra as a tool for supporting reasoning and consequently they push teachers towards a change of their practices as a result of the discussions they have in their schools about the nature of the highly important SVN tests.

As pointed out above, this type of item is an important stimulus for reflection by teachers, to consider a new approach to the culture of theorems at school, challenging normal teaching practices. Usually in Italy (and possibly also in other countries) the teacher asks the students to understand and repeat proofs of statements he or she has supplied, rather than prove statements. Even more seldom students are asked to produce conjectures themselves or to justify a statement.

The aim of this type of item is to change teaching practices in the school, harnessing the strong impact that the SNV tests have on teachers' practices. In fact proving activities are not generally common in the first years of Italian secondary schools, particularly using any algebraic machinery. Most practices in algebra in Grades 9 and 10 are more concerned with the manipulative aspects of formulas and not its use as a thinking tool that can support mathematical reasoning (Arzarello et al. 2001). This appears only later and only in some the more scientifically oriented schools with a stronger mathematics curriculum, when elementary calculus is introduced.

Discussion

In this chapter we have illustrated how the debate originating from the disappointing results of Italian students in the 2003 PISA study had a positive impact in the country. First, it convinced the Ministry of Education to design a national policy for assessing the quality of teaching in the schools by establishing an Italian Assessment System (SNV). It gradually started a systematic annual census survey at selected grades. Second, the Ministry promoted seminars about the meaning of PISA studies and innovative programs for the teaching of mathematics, which involved a considerable number of Italian mathematics teachers.

We have also illustrated how the SNV framework is strongly but not completely aligned with that of PISA. A feature of the Italian items, which distinguishes them from those of PISA, is the presence of items where students are asked about arguing and proving in completely intra-mathematical contexts. This is an aspect of mathematics that is not part of the OECD's defined mathematical literacy, but is due to cultural instances that feature within the Italian curriculum and to the consequent necessity of testing such competencies.

All these PISA-driven initiatives in Italy are having a positive influence on the results of the most recent international assessment studies. For example the recently published results of PISA 2012 (OECD 2014a) confirm a positive trend for Italy, even though the results are still below the OECD mean (mean score 485, SE 2.0). In particular they confirm the 2009 improvement for southern regions (see Fig. 13.2 above), even though the differences between the southern and northern regions remain high.

A wide-ranging study for the reasons of this remarkable change has not yet been carried out, but, based on our experience and knowledge of what happens in schools, we provide here a tentative explanation of this phenomenon:

(i) The gradual introduction of SNV from 2008 has called to teachers' attention the meaning of standardised international and national assessment systems. At the beginning, programs to measure reading and mathematical literacy were almost unnoticed by the majority; but in a short period, school communities became strongly focussed on them.

(ii) The SNV activities are carried up each year in May for Grades 2, 5, 6 (not from 2014), 10, and in June for Grade 8. In July the whole country receives a picture of the macro-situation of Italian schools, since results drawn from a sample of schools are made public. In October each school knows its own results. This causes a careful and serious reflection by people working in the school (teachers, principals, regional and national school officers) and outside it (families, policy makers). The discussion has shifted from the acceptance or non-acceptance of the standardised national survey to the relationships between accountability and improvement (Hargreaves and Braun 2013). As a consequence also the international surveys, and especially PISA, are considered and compared with the results of the SNV.

(iii) A further element of synergy between the international and national surveys is that most of the students who participated to PISA 2012 had also participated in the Grade 10 SNV survey of that year, and in 2010 had participated in the Grade 8 national survey, which was part of their final examination at the end of the first cycle of instruction. For the first time in Italy it has been possible to compare two standardised surveys for a comparable group of students (Montanaro 2013). Even though the two surveys have different aims and frameworks, they have started a useful discussion.

(iv) The programs for updating teachers about the SNV and PISA surveys, promoted by the Ministry of Education, point out more and more the similarities and differences between the two. Consequently, teachers' attention has gradually shifted its focus from the overall results to the analysis of their items and to the scrutiny of the frameworks behind them. There has been a shift in concern from “What are the results?” to “How have students responded and why?” This change in perspective seems to be confirmed by the survey about teachers’ use of cognitive activation strategies (OECD 2014b), where Italy’s results on the constructed index are near the OECD mean (−0.10, SE 0.02). This indicates that teachers are giving a certain attention to students’ thinking processes. PISA reports the estimated increase in mathematical literacy scores for each unit increase in this index and for Italy it amounts to 11.3 which is one of highest among OECD countries.

For all these reasons, signs of a positive change are on the horizon.

References

Anichini, G., Arzarello, F., Ciarrapico, L., & Robutti, O. (Eds.). (2004). *New mathematical standards for the school from 5 through 18 years, the curriculum of mathematics from 6 to 19 years, on behalf of UMI-CIIM, MIUR (edition for ICME 10)*. Bologna: UMI.

Arzarello, F., Bazzini, L., & Chiappini, G. (2001). A model for analysing algebraic processes of thinking. In R. Sutherland, T. Rojano, A. Bell, & R. Lins (Eds.), *Perspectives on school algebra* (pp. 61–81). Dordrecht: Kluwer.

Arzarello, F., Bernardi, C., Borgi, R., Ciarrapico, L., De Santis, F., Naldini, M., et al. (Eds.) (2012). m@t.abel mathematics for students on the threshold of the third millennium. http://mediarepository.indire.it/iko/uploads/allegati/M7PWITOE.pdf. Accessed 23 Aug 2013.

Boero, P., Garuti, R., & Lemut, E. (2007). Approaching theorems in grade VIII. In P. Boero (Ed.), *Theorems in school* (pp. 250–264). Rotterdam: Sense.

Garuti, R., & Boero, P. (1994). Mathematical modelling of the elongation of a spring: Given a double length spring. In J. P. da Ponte, & J. F. Matos (Eds.), *Proceedings of the eighteenth international conference for the psychology of mathematics education (PME XVIII)* (vol II, pp. 384–391). Lisbon: Department of Education, University of Lisbon.

Hargreaves, A., & Braun H. (2013). Data-driven improvement and accountability. Boston College, National Education Policy Center. http://nepc.colorado.edu/files/pb-lb-ddiapolicy.pdf. Accessed 13 May 2014.

INVALSI. (2011). Gli esiti del Servizio nazionale di valutazione 2011 e della prova nazionale 2011. http://www.invalsi.it/snv1011/documenti/Rapporto_SNV%202010-11_e_Prova_nazionale_2011.pdf. Accessed 23 Aug 2013.

Mariotti, M. A. (2006). Proof and proving in mathematics education. In A. Gutierrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education* (pp. 173–204). Rotterdam: Sense.

Montanaro, P. (2013). Un confronto tra Pisa 2012 e le rilevazioni nazionali 2011–12. http://www.invalsi.it/invalsi/ri/pisa2012/rappnaz/pres/Confronto_PISA-RN.pdf. Accessed 14 May 2014.

Organisation for Economic Co-operation and Development (OECD) (2004). The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills, PISA, OECD Publishing. doi:10.1787/9789264101739-en

Organisation for Economic Co-operation and Development (OECD) (2010). PISA 2009 results: Learning trends: Changes in student performance since 2000 (vol. V). http://dx.doi.org/10.1787/9789264091580-en. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD) (2013). PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy, PISA, OECD Publishing. doi:10.1787/9789264190511-en

Organisation for Economic Co-operation and Development (OECD) (2014a). PISA 2012 results: What students know and can do: Student performance in mathematics, reading and science (vol. I). http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf. Accessed 13 May 2014.

Organisation for Economic Co-operation and Development (OECD) (2014b). PISA 2012 results: Ready to learn: students' engagement, drive, and self-beliefs (vol. III). http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-iii.htm. Accessed 14 May 2014.

Chapter 14
The Effects of PISA in Taiwan: Contemporary Assessment Reform

Kai-Lin Yang and Fou-Lai Lin

In Taiwan, PISA used to be an inactive seed
Appears once every three years
Could only be seen in newspapers.
Taiwanese performance in PISA
Seemed to be similar in TIMSS.
Be excellent in mathematics and science literacy
But poor in reading literacy.
After summer 2012, the inactive seed suddenly burst
Into every family with high school students,
Into the minds of all high school teachers,
Into daily conversations of Taiwanese educational community.
Meanwhile, a strange phenomenon arose:
PISA cram schools shot out numerously.
This chapter aims to report the dramatic effects
And investigate the reasons behind.

Abstract Taiwan has always been one of the top ranked countries in PISA, so initially interest in PISA was mainly concerned with standards monitoring, with some analysis of how instruction could be improved. However, from 2012, PISA became a major public phenomenon as it became linked with proposed new school assessment and competitive entrance to desirable schools. Students, along with their parents and teachers, worried about the ability to solve PISA-like problems and private educational providers offered additional tutoring. This chapter reports and explains these dramatic effects. Increasingly, the PISA concept of mathematical literacy has been used, along with other frameworks, as the theoretical background for thinking about future directions for teaching and assessment in schools. This is seen as part of an endeavour to change the strong emphasis on memorisation and repetitive practice in Taiwanese schools.

K.-L. Yang (✉) • F.-L. Lin
Department of Mathematics, National Taiwan Normal University,
No. 88 Sec 4, Ting-Chou Rd., Taipei, Taiwan
e-mail: kailin@ntnu.edu.tw; linfl@math.ntnu.edu.tw

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_14

Background to the Taiwanese Educational System

Before 1967 only elementary school education, for children from 7 to 12 years old, was compulsory in Taiwan. In order to gain entry to a favoured junior high school, many 11 and 12 year old students attended after-school classes, often colloquially called the cram-schools. In Taiwanese, these are called *buxiban*. In order to release students from the pressures of competitive entry to limited places in junior high schools and thereby postpone the time for attending buxiban, in 1968 compulsory education was extended to 3-year junior high schools (12–15 years old). Besides, the public believed that the length of compulsory education reflects the modernity of a country. Therefore, some politicians extended compulsory education to attract votes. Now, because of the educational policies of our current president, Ying-jeou Ma, a 12-year compulsory education program that integrates primary education, junior high, senior high or vocational education will be implemented in 2014.

The Taiwanese Mathematics curriculum for the years of compulsory education has undergone three reforms in the last three decades. Based on the shift of focus towards students as knowledge constructors, the first reform was to revise the 1975 Standards of School Mathematics Curriculum to emphasise the manipulation of concrete materials. Nevertheless, the revised Standard of School Mathematics Curriculum in 1993 resulted in having the two systems of algorithmic mathematics and mathematics with manipulatives coexisting in classrooms: the formal taught methods alongside the child-invented methods, as described by Booth (1981). In order to complement the defects of previous curriculum, further reforms were still required.

The second and third reforms included the 2000 Nine-Year School Curriculum, issued in 1998 and then revised in 2003. The 2000 Nine-Year School Curriculum was proposed with a basic philosophy of constructivism and it emphasised the value of children's own methods. However, after the implementation of this new curriculum in 2002, the seventh grade students did not perform well on the first mathematics examination. Thus some scholars, especially the mathematicians, asked the Ministry of Education to revise the 2000 mathematics curriculum and a reform group was formed (Leung et al. 2012). The major differences between the Nine-Year School Mathematics Curriculum of 2000 and 2003 lay in the quantity and sequence of the content. The 2000 Mathematics Curriculum expected 80 % students to keep up with the scheduled content, so the content was less, simpler, and flexibly divided into four learning stages. However, the 2003 Mathematics Curriculum presupposed that no more than 50 % students would be left behind the scheduled content. Consequently, the content sequences for each year were listed according to ability indicators. However, the content was more inflexible and more difficult than that of 2000.

Buxiban are private schools that offer out-of-school instruction to improve students' achievement scores. In Taiwan, buxiban are so popular that they have become a sunrise industry. The major function of these buxiban is to increase the possibilities of getting into desirable high schools and universities. These schools mainly focus on enhancing students' academic abilities in mathematics, science, English, and Chinese writing. For Taiwanese students of Grade 11, it has been found that attending a buxiban improves educational achievement, and 49 % of eleventh graders reported they spent several hours each week in buxiban (Chen and Lu 2009). The main reasons why students go to buxiban include (1) following established customs of going to buxiban, (2) as a way to make friends, (3) their fears of getting academically behind classmates who go to buxiban, and (4) unsatisfactory performance in school examinations.

Taiwanese Students' Performance in PISA

Taiwan has participated in PISA since 2006. In mathematics, it was ranked 1 in 2006 (mean score 549) and 5 in 2009 (mean score 543). Taiwanese students were relatively better in mathematics and science (ranked 12 in 2009) than in reading (ranked 23 in 2009) although the correlations between the three scores at the level of the student are very high (0.81–0.85). Although Taiwanese students' performance in mathematics and science literacy is internationally ranked at the top level (OECD 2009, 2012), the achievement gaps between high and low achievers are larger than in many other countries, so this is something that needs attention. Table 14.1 shows the percentage of students at each level of mathematical literacy in Taiwan and the other countries ranked in the top six for PISA 2009, and also for Taiwan in PISA 2006. The percentage of students at and below level 1 is larger for Taiwan than the other high-performing countries. First results from PISA 2012 indicate that this pattern continues.

Although Taiwanese students' average performance in mathematical literacy is internationally top-ranked, it is still profitable to study their weaker areas to guide further improvement. By analysing Taiwanese students' responses, several places where improvements might be made have been identified. Firstly, some students

Table 14.1 The percentage of students from high performing countries at different levels of mathematical literacy in PISA 2009

Nation	Country rank	% below 1	% at level 1	% at levels 2, 3, 4	% at level 5	% at level 6
Taiwan	5	4.2	8.6	58.6	17.2	11.3
Taiwan (2006)	1	3.6	8.3	56.1	20.1	11.8
Finland	6	1.7	6.1	70.5	16.7	4.9
Korea	4	1.9	6.2	66.3	17.7	7.8
Shanghai	1	1.4	3.4	44.7	23.8	26.6
Hong Kong	3	2.6	6.2	60.5	19.9	10.8
Singapore	2	3.0	6.8	54.6	20.0	15.6

were not familiar with connecting given situations and their descriptions with figures to make reasonable assumptions, e.g. statistical graphs. Secondly, although the mathematical models behind problems situated in real-world contexts were not hard for students, some students were distracted by superfluous but related information in a problem. This implied that they were relatively weak in discriminating relevant and irrelevant information to solve problems in real-world contexts. Thirdly, some students did not correctly answer estimation problems, which may result from lack of familiarity with estimating large numbers, computing with calculators, or thinking of tolerable errors. All of this may arise because Taiwanese students are often 'stuffed with a standardised answer'. Fourthly, some students tended to provide personal interpretations rather than evidence-based explanations and then their over-inference caused wrong answers. This showed that they did not understand that valid information would be the basis of strong explanations. The weaknesses of Taiwanese classroom teaching were revealed by the above-mentioned features and partially resulted from the strategies teachers used. Due to the prevalence of multiple-choice tests, many Taiwanese teachers teach students strategies of deletion and substitution, which can be used for quickly isolating a correct answer.

Two Literacies for Selecting High Achievers

Taiwan is going to implement a 12-year compulsory education program in 2014. In general, senior high schools, being compulsory, should have open admission for junior high school students if they meet the minimum test score and other relevant requirements. However, the reality is more complex. Currently, senior high schools are ranked hierarchically according to their students' entrances scores in the national examination. The most desirable high schools have the highest scores. Consequently, the top 15 % of students, in particular, are nearly all gathered into specific schools. It is not yet certain whether this will continue in the future, or whether they will be spread among many different schools that all may offer a special curriculum for high achievers.

Although there is an examination to evaluate the competency of students in Grade 9, the main assessment goal has traditionally been students' mastery of textbook content rather than their 'learning power', which is more important. Learning power is based on thinking and reading, and therefore mathematical literacy and reading literacy should be assessed. In this context, mathematical literacy is defined as using mathematical knowledge and skills to identify and solve situational or mathematical problems, and understanding written text to reflect on mathematical knowledge included in the Taiwanese curriculum. It is a new challenge to identify the high achievers in the top 15 % according to their learning power rather than just their content knowledge. This definition of mathematical literacy has been inspired by the PISA definition and adapted to suit the purposes of assessment in Taiwan.

Alternative Assessment Goals and Framework

When we acknowledge that "mandated assessment mediates between the expectations of the system and their embodiment in classroom practices" (Barnes et al. 2000, p. 626), we come to realise that the alternative assessment goals for high achievers should be a tool to reshape school practices. Instead of considering only a selection function, they should also consider the key purposes of teaching and learning in compulsory education.

Accordingly, Lin (2012) analysed the consequences of deciding that a major educational goal was to enhance students' learning power. He elaborates learning power in three dimensions: tools, learning methods and dispositions. The three dimensions of learning power support an analytical approach to assessment reform. Language and thinking are two necessary tools, while reading and inquiry are two main learning methods. Dispositions refer to learners' emotions, attitudes, and beliefs. This is in accordance with the definition of learning power as

> a complex mix of dispositions, lived experiences, social relations, values, attitudes and beliefs that coalesce to shape the nature of an individual's engagement with any particular learning opportunity. (Deakin Crick et al. 2004, p. 247)

Both language and thinking are required in learning different subjects. On the one hand language, especially as reading literacy, is an interdisciplinary competency. On the other hand, mathematical literacy supports logical thinking and forms the basis for pursuing advanced knowledge. Therefore, the assessment goals are to measure mathematical literacy (how students use the knowledge and skills they have acquired at school to solve open-ended and reasoning problems) and reading literacy (how to gain knowledge from reading text in multiple disciplines including history, geography, civics and science). Students' dispositions are not included in this assessment reform because they cannot easily be objectively evaluated and ranked through a time-limited, paper-and-pencil test.

For the proposed assessment of mathematical literacy, we adopted three components from the PISA Mathematics Framework. The first component was mathematical content organised around overarching ideas such as *Quantity*, *Space and shape*, *Change and relationships*, and *Uncertainty* (OECD 2004). The second component was the use of context so that problems are set in various real-world situations. The third component was mathematical competencies. The mathematical competencies were considered to be more critical than the other two components in order to discriminate the level of mathematical literacy of high achievers. For this assessment, the most important competencies are problem solving, reasoning and proof. (Note that this notion of competencies draws on but is not the same as that described in Chap. 2 in this volume). Taiwanese students' performance in PISA placed about 30 % of students at levels 5 and 6 of mathematical literacy. The feature of these two levels, as described in the Mathematics Framework (Taiwan PISA National Center 2011) is being able to handle complex problems and advanced reasoning. In order to provide an assessment for selection of the very best students, we only focused on problem solving and reasoning and proof, which also correspond to the top level of mathematical competence delineated by Jan de Lange (1999).

Feasibility of this Framework

The feasibility of this framework will be verified by its relevance to the Taiwanese School Mathematics Curriculum and the empirical validity of selecting high-achievers. Ideally, assessment tasks should match the expectations of curriculum documents, syllabuses, or courses of study. Although the framework above is not directly related to the national curriculum, two components of the framework, the overarching ideas (now called content categories as in OECD 2013) and the mathematical competencies fit the spirit of the curriculum. To be more specific, the national curriculum aspires to connect different mathematics units to different learning domains inside and outside of mathematics, applying mathematics to daily life, appreciating the beauty of mathematics, and further cultivating interest in exploring the essence of mathematics as well as other related disciplines (Ministry of Education 2003). The PISA content categories are not directly drawn on but are in reasonable correspondence with the Taiwanese School Mathematics Curriculum. The mathematical competencies just match its spirit. Only the mathematical models underlying task situations that are directly within the Taiwanese School Mathematics Curriculum will be included in the PISA-like assessment. This is so that it can assess all high-achievers equitably. Even though the mathematical content of the reformed assessment is constrained by the national school curriculum, the scope of mathematics is deeper and broader than PISA in order that it can validly select the top-achievers.

Before proposing this framework, Lin (2011) invited 25 mathematics educators to help 180 junior high school teachers understand mathematical literacy from the perspective of PISA and also using the ideas on mathematical proficiency expressed in the book “Adding It Up” from the United States (National Research Council 2001). The first aim of this collaboration was to help teachers clarify the difference between mathematics for the promotion of mathematical literacy and the mathematics of examination; a distinction that should be well-known by teachers but might be easily confused. The features behind the mathematics of examination include closed problems, one problem with one predetermined answer, precise information, and problems posed in decorative (but not necessarily realistic) situations. On the contrary, the features behind mathematics to promote mathematical literacy included open-ended problems; problems with more than one possible approach; problems with multiple plausible answers; problems with superfluous information and ‘productive situations’. Productive situations give clues for students to connect a situation with various related mathematical concepts and then to produce multiple assumptions or a required transformation between situational and mathematical worlds.

During the workshop for designing assessment tasks for mathematical literacy, Lin posed several problems to enhance teachers' perception of the difference between mathematics of literacy and mathematics of examination. For example, he posed the question of estimating the lowest threshold to win an election if eight representatives are to be chosen from 200 members. In general, teachers automatically answered this question based on the assumption that one member only had one vote. The discussion of this problem was used to highlight the fact that assumptions about situations could be implicit and multiple, and that different answers were plausible depending on the assumptions made.

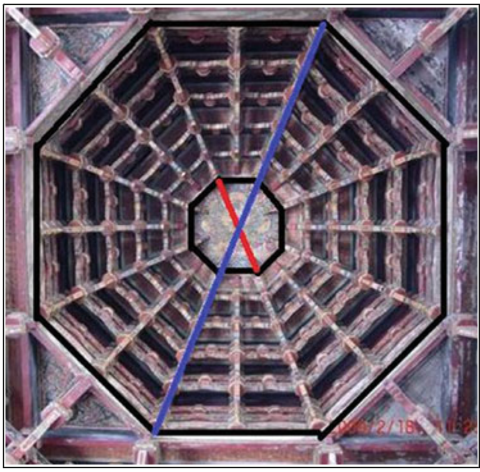
Then, they cooperatively designed about 180 problems that aimed at assessing students' abilities of using conceptual understanding, procedural fluency, strategic competence, and adaptive reasoning (the elements of mathematical proficiency from the report of Kilpatrick et al. (National Research Council 2001)) to formulate mathematical models, provide mathematical answers, explain mathematical answers in situations, and critique mathematical models or answers. The scope of the 180 problems differs in several ways to the set of PISA problems. One difference is that they deliberately connect different mathematical units, for example the distance between two points at rectangular coordinates and the Pythagorean Theorem. To score each question, PISA uses at most three levels (0, 1, 2) whereas the PISA-like assessment scores are classified into more levels because the mathematics content is much more complex than PISA. Moreover, although the scoring rubric for the assessment is precisely described, it is still a challenge to get consistent assessment across a large number of teachers.

Another consideration is that mathematical proof is not specifically included in the PISA items (given its focus derived from the OECD mandate for life skills), but needs to be included in the reformed assessment in Taiwan. In mathematics, proof is the rigour and logical connection among mathematical knowledge and this greatly differs from proofs outside of mathematics. Assessing proof is essential when the purpose is not only to select the top 15 % of achievers but also to rank them. Mathematical proof is a special text genre in written discourse (see Pimm and Wagner 2003), and the ability to read a mathematical proof requires both mathematical knowledge and deductive reasoning (Lin and Yang 2007). On the contrary, but in accordance with its definition of mathematical literacy, PISA mainly considers plausible reasoning. In Taiwan, a research study showed that 5.7 % of Grade 9 students could give complete arguments to prove that 'the sum of any two odd numbers must be even' and a further 37.2 % could give partial arguments that included all information but omitted some reasoning (Lin et al. 2004). Around 36 % of Grade 9 students could construct a correct proof, which required combining several geometric arguments (Heinze et al. 2004) and 18.8 % of Grade 9 students were scored in the top level of reading comprehension of geometric proof (Lin and Yang 2007). Thus, in the Taiwanese situation, items requiring constructing or comprehending proof were also considered to be a necessary and viable part of assessing mathematical literacy of the highest achievers.

Problems Exemplifying Mathematical Literacy

In this section, we provide three problems to exemplify the assessment of mathematical literacy for high achievers. Figure 14.1 shows a problem about the geometry underlying antique architecture. Students are required to actively use rulers to figure out the scale of this picture and then to estimate the length of the diagonal line in the innermost layer of the octagon. An adequate solution is to measure the length of the long diagonal line in Fig. 14.1 with a ruler, then calculate the proportional scale using the known measurement of 5.5 ft. Measure the length of the short diagonal line, then calculate the real length using the scale. Figure 14.2 shows an uncertainty problem concerned with data about buxiban students. Students need to actively identify one advantage with regard to each buxiban and represent this advantage with a suitable statistical chart. For one buxiban, the pass rate (as a percentage) is the highest, for another the absolute number of students passing is the highest, and the third shows steady improvement. Figure 14.3 shows a paper-folding problem where students need to prove the obtained triangles are equilateral.

The content of these questions is included in the Taiwanese junior high school curriculum, and the situations come from students' life experiences. Nonetheless, our students are unfamiliar with these kinds of questions due to the need to identify that some information is superfluous, the need for to make assumptions and the openness of the potential problem solving strategies.



Here is the Eight Trigrams shaped ceiling of Lu-Gang Longshan Temple, the biggest ceiling in Taiwan. Its span, the diagonal line shown over the outermost layer of the octagon, is about 5.5 feet and the height of the top centre is about 6.5 feet. It is tiered up with five layers, each made of 16 crossbeams to support the weight of the roof eaves. The crossbeams are carved with exquisite sculptures from Chinese culture. The ceiling is built using nails of wood rather than metal. The ceiling is filled with the wide and deep wisdom of our ancestors.

Please estimate the length of the diagonal line in the innermost layer of the octagon.

(1 foot = 12 inches, 1 inch = 2.54 cm)

Fig 14.1 Eight Trigrams shaped ceiling and a problem for Grade 9 students

There are three competitive buxiban in Ting-Sou's hometown. The number of students attending the buxiban and the number of these passing the Basic Competence Test for Junior High School Students are shown in the table for the past three years Because of the keen competition among these three buxiban, if they ever use any false data, that buxiban will be attacked by the other two. Consequently, the buxiban will lose its credit and students, and have to pay a fine for false advertisement. Therefore, all the buxiban use real data to design favourable flyers for themselves. Please answer the following questions using the data in the table.

If you are the publicity manager of one of these buxiban, how would you design a statistical chart to highlight the advantage of your company? (Provide an answer for each buxiban.)

Buxiban	96th academic year		97th academic year		98th academic year	
	Number of students	Number of passes	Number of students	Number of passes	Number of students	Number of passes
P	60	30	65	31	80	32
S	130	40	120	40	125	40
Q	35	12	42	15	39	15

Fig. 14.2 Buxiban advertisement problem

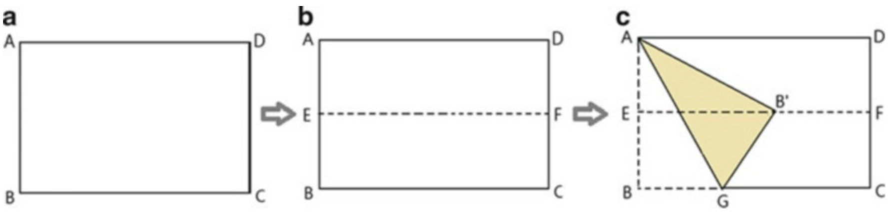


Fig. 14.3 Paper-folding problem

PISA: Insanity and Retreat

PISA-like assessment reform has become a storm in Taiwan. Three different types of 'PISA insanity' are illustrated by the news report in Fig. 14.4, which has been translated by the authors. The news item shows that the effects of PISA-like assessment are found on parents, on buxiban, and on governmental policies. Parents are worried about their children's failure in the entrance examination. Buxiban are sensitive to the disturbance and take advantage of the assessment reform to make money. Whether there is any positive effect on students' learning is still questionable. The government is advancing several programs for high school teachers to

PISA Assessment for Entrance Examination in Keelung and Greater Taipei: Parents Are Much More Worried than Students

- The Chairman of the Secondary School Parents Association in Taipei, Mr Young-Jia Hsu, has criticised some buxiban that take advantage of the panic and anxiety of parents and students to recruit students into PISA training sessions. No matter what the effects are, this situation is similar to fraud.
- The reporter visited several buxiban with PISA training sessions in Taipei and found the cost is about NTD$650-850 per lesson, which is up to10% higher than general courses at the buxiban.
- The Deputy Chief of Department of Education, Taipei City Government, Dr Ching-Huang Feng states that the Comprehensive Assessment Program for Junior High School Students includes the traditional five courses in the assessment. If the special high school admission examination takes the same courses as well, Taiwanese secondary teaching will follow our old route only emphasising memorisation and repeated practice. However, the Keelung and Greater Taipei regions will include literacy courses in the assessment as has been publicised widely. Without this, it would be difficult to make any change. Therefore, the Department of Education, Taipei City Government has requested schools to include literacy questions in general assessments for Grade 7 and 8 students, in the hope that students will gradually become familiar with questions of this kind, and so be confident when participating in the special high school admission examination.

Fig. 14.4 Article from the China Times of 1 January 2013 (Lin et al. 2013)

better understand the PISA-like assessment and to design tasks for developing students' mathematical literacy. In addition, it is suggested that PISA-like problems should be included in each regular test. Some teachers agree with this reform but others do not. The voices querying the reform are continuously represented in mass media by professors, teacher representatives, parent representatives, and ordinary people; in particular, some of them expressed concern that different scoring criteria would result in unfairness. As to the traditional examinations, the scoring codes referred to one standardised answer with several key steps. The more different students' answers and the key steps are, the lower scores would be obtained. In PISA-like assessment, the scoring codes refer to multiple plausible answers. The more plausible answers completed, the higher scores would be given until full marks. That is to say: traditional examinations mainly tested what students had not comprehended, but PISA-like assessment focuses on what students should have learned. As a consequence, mathematics teachers may spend more time discussing their ideas about mathematics, its learning and teaching with each other.

Several months ago, we were all preparing for the PISA-like assessment. However, in the Taipei City Council in April 2013, regional delegates rejected the proposed assessment reform based on mathematical literacy and reading literacy for selecting the top students, and we suppose there may be progressive transition to the PISA-like assessment. There were multiple factors in the opposition. Most people felt the move was too hasty and there had been inadequate support. When news of this assessment reform was publicised, the database of sample questions was embryonic, and the scoring criteria and the exact time for executing the assessment were still uncertain. Some people oppose the whole idea of selection to the 'star schools' and think students should attend local schools in the compulsory years of education. As students concentrate on only mathematics and reading literacy, teachers of other subjects in the buxiban have fewer students and so oppose the reform, and even teachers of the newly assessed subjects are against it because it does not match their regular teaching. The last straw was doubt about the fairness of the reformed assessment. Taiwanese are used to being 'force-stuffed' with a standardised answer to a standardised problem; we did not believe open-ended problems could make a fair assessment. In a diploma-driven traditional society like Taiwan, there is always great public concern about assessment.

Reflection

After reflecting on the failure of the assessment, we agree that it is important to align assessment with classroom teaching and learning. However, the premise for such incorporation lies in whether classroom teaching and learning are appropriate to enhance students' learning power. Based on the fate of this reform, we confess that the national assessment reform was not supported across the system. Failure demonstrated that sometimes political issues are much more influential than the assessment, teaching and learning. Assuming optimistically that the PISA-like assessment has been postponed rather than cancelled, our preparations for it continue. For example, the National Academy for Educational Research, the Curriculum & Instruction Consulting Team of Ministry of Education, and the National Science Council will continue with projects to develop students' mathematical literacy. Through longer term projects of developing and implementing a new educational system in harmony with the goals of assessment reform, we believe the concern about reform will be eased, the beliefs about the fairness of non-traditional assessment will build up, and the uncertainty surrounding the new assessments will be eliminated. We are also confident that the emergence of the alternative assessments will be beneficial to improve and not undermine our classroom teaching and learning. Hopefully, after a further 3 years of effort in preparation for PISA-like assessment, it will be successfully implemented and it will stimulate our education to focus on developing students' learning power rather than teaching just for examinations.

In 2009, Grek wrote

> The construction of PISA with its promotion of orientations to applied and lifelong learning has powerful effects on curricula and pedagogy in participating nations, and promotes the responsible individual and self-regulated subject. (Grek 2009, p. 35)

She noted that PISA data were applied to justify changes or provide support for domestic and European policy-making processes to different extents in different countries: from the PISA-promotion of the UK, the PISA-shock of Germany to the PISA-surprise of Finland. Like these Western European countries, Taiwan is experiencing the effects of PISA. In the past, PISA data was applied to check whether Taiwanese students were retaining their top ranking. Now, PISA's theoretical background and assessment Framework strongly influence thinking about examinations and teaching in schools.

References

Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: The engine of systematic curriculum reform. *Journal of Curriculum Studies, 32*(5), 623–650.

Booth, L. R. (1981). Child-methods in secondary mathematics. *Educational Studies in Mathematics, 12*(1), 29–41.

Chen, S. Y., & Lu, L. (2009). After-school time use in Taiwan: Effects on educational achievement and well-being. *Adolescence, 44*(176), 891–909.

de Lange, J. (1999). *Framework for classroom assessment in mathematics*. Utrecht: National Centre for Improving Student Learning and Achievement in Mathematics and Science.

Deakin, C. R., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education, 11*(3), 247–272.

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy, 24*(1), 23–37.

Heinze, A., Cheng, Y. H., & Yang, K. L. (2004). Students' performance in reasoning and proof in Taiwan and Germany: Results, paradoxes and open questions. *Zentralblatt fur Didaktik der Mathematik, 36*(5), 162–171.

Leung, S. K., Yang, D. C., & Leu, Y. C. (2012). Taiwan mathematics curriculum, its historical development and research studies. In T. Y. Tso (Ed.), *Proceedings of the 36th conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 165–178). Taipei: PME.

Lin, F. L. (2011). *Sample items developed in Taiwan for assessing mathematical literacy* (in Chinese). http://pisa.nutn.edu.tw/sample_tw.htm. Accessed 14 Oct 2013.

Lin, F. L. (2012). Proposing the new policy by using 'Reading and Thinking' to cultivate students' learning power according to the current educational issues in Taiwan. Educational Forum, *Foundation of Commonwealth Magazine*. Taipei/Changhua.

Lin, F. L., & Yang, K. L. (2007). The reading comprehension of geometric proofs: The contribution of knowledge and reasoning. *International Journal of Science and Mathematics Education, 5*(4), 729–754.

Lin, F. L., Yang, K. L., & Chen, C. Y. (2004). The features and relationships of explanation, understanding proof and reasoning in number pattern. *International Journal of Science and Mathematics Education, 2*(2), 227–256.

Lin, C. C., Hu, C. H., & Chu, F. Y. (2013). PISA assessment for entrance examination in Keelung and Greater Taipei: Parents are much more worried than students. *China Times*. http://news.chinatimes.com/focus/501012570/112013010100354.html. Accessed 23 Aug 2013 (in Chinese).

Ministry of Education. (2003). *Grade 1–9 mathematics curriculum formal guideline in Taiwan*. Taiwan: Ministry of Education (in Chinese).

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.

Organisation for Economic Co-operation and Development (OECD). (2004). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: PISA, OECD Publishing. doi:10.1787/9789264101739-en.

Organisation for Economic Co-operation and Development (OECD). (2009). *PISA 2006 technical report*. Paris: OECD.

Organisation for Economic Co-operation and Development (OECD). (2012). *PISA 2009 technical report*. Paris: OECD Publishing. Retrieved March 4, 2013, from http://dx.doi.org/10.1787/9789264167872-en

Organisation for Economic Co-operation and Development (OECD). (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. doi:10.1787/9789264190511-en.

Pimm, D., & Wagner, D. (2003). Investigation, mathematics education and genre: An essay review of Candia Morgan's writing mathematically: The discourse of investigation. *Educational Studies in Mathematics, 53*(2), 159–178.

Taiwan PISA National Center. (2011). *Taiwan PISA 2009 report*. Taiwan: Taiwan PISA National Center (in Chinese).

Chapter 15
PISA's Influence on Thought and Action in Mathematics Education

Kaye Stacey, Felipe Almuna, Rosa M. Caraballo, Jean-François Chesné, Sol Garfunkel, Zahra Gooya, Berinderjeet Kaur, Lena Lindenskov, José Luis Lupiáñez, Kyung Mee Park, Hannah Perl, Abolfazl Rafiepour, Luis Rico, Franck Salles, and Zulkardi Zulkardi

Abstract This chapter contains short descriptions from contributors in ten countries (Chile, Denmark, France, Indonesia, Iran, Israel, Korea, Singapore, Spain and USA) about some ways in which the PISA Framework and results have influenced

K. Stacey (✉) • F. Almuna
Melbourne Graduate School of Education, The University of Melbourne, 234 Queensberry Street, Melbourne, VIC 3010, Australia
e-mail: k.stacey@unimelb.edu.au; felipea@student.unimelb.edu.au

R.M. Caraballo • J.L. Lupiáñez • L. Rico
University of Granada, Granada, Spain
e-mail: caraba@correo.ugr.es; lupi@ugr.es; lrico@ugr.es

J.-F. Chesné
Office for the Evaluation of Educational Activities and Experimentations, Ministry of National Education, DEPP, 65 rue Dutot, Paris, France
e-mail: jean-francois.chesne@education.gouv.fr

S. Garfunkel
Consortium for Mathematics and its Applications, Bedford, MA, USA
e-mail: s.garfunkel@comap.com

Z. Gooya
Shahid Beheshti University, Tehran, Iran
e-mail: zahra_gooya@yahoo.com

B. Kaur
National Institute of Education, NIE 7-03-46, 1 Nanyang Walk, Singapore 637616, Singapore
e-mail: berinderjeet.kaur@nie.edu.sg

L. Lindenskov
Institut for Uddannelse og Pædagogik, Aarhus Universitet i Emdrup, Copenhagen, Denmark
e-mail: lenali@dpu.dk

K.M. Park
College of Education, Hongik University, Seoul, South Korea
e-mail: kpark@hongik.ac.kr

H. Perl
Ministry of Education, 2 Dvora Hanevia St., Jerusalem 9100201, Israel
e-mail: hannahpe@education.gov.il

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7_15

thinking and action about mathematics education. In many countries, the PISA results have been a call to action, and have stimulated diverse projects aimed at improving results, principally for teacher education but also some involving students. PISA resources, including the released items, have been used as a basis for assessment as well as for teacher development. Some countries have established national assessments with noticeable consistency with PISA ideas. In many countries, PISA's concept of mathematical literacy, with its analysis of what makes mathematics education useful for most future citizens, has been extremely influential in curriculum review and also for improving teaching and learning. Countries have also incorporated or adopted the way that PISA describes mathematical competence through the fundamental mathematical capabilities.

Introduction

The aim of this chapter is to review some of the ways in which PISA has influenced thinking about mathematics education in a variety of countries around the world, and to document some of the actions that have followed from this influence. The chapter consists of ten separate, short pieces that are contributed by citizens of various countries. This collection is designed to complement the more substantial contributions from Germany, Italy, Japan, and Taiwan in the earlier chapters of this volume. Invitations to contribute to this chapter were issued to people who were likely to be in a position to make a sound judgement, sometimes because of their involvement with the national team implementing PISA or sometimes because of their long term involvement with curriculum and teaching issues more generally or for their other special interest. However, these are generally personal pieces and do not represent all the action or opinions in a country, nor are they definitive evaluations of the local influence of PISA. Instead they are personal reflections (some more so than others) written from the point of view of people involved in various ways with the national agendas. To assist in interpretation, the sections begin with a very brief description of the contributors' local roles. Contributions are presented alphabetically by country name.

A. Rafiepour
Shahid Bahonar University of Kerman, Kerman, Iran
e-mail: drafiepour@gmail.com

F. Salles
Office for Students' Assessment, Ministry of National Education,
DEPP, 65 rue Dutot, Paris, France
e-mail: franck.salles@education.gouv.fr

Z. Zulkardi
Sriwijaya University, Palembang, Indonesia
e-mail: zulkardi@yahoo.com

In drawing conclusions about the extent of influence of PISA around the world, it is important for readers to know that contributions were not solicited or selected from countries where PISA was known to have been especially influential. It is also relevant that everyone who was invited to contribute had something to report about their country.

When the contributions are reviewed as a set, it is evident that PISA has had a substantial influence on both thought and action in many countries. The country ranking and the mean scores of students and their distribution have been important, sometimes to affirm national directions as in the case of Singapore, but more often as a stimulus to action especially where student performance has been lower than expected. The type of action taken is varied. In some countries, including Spain, international assessment has been supplemented by new forms of national assessment, sometimes based around a PISA-like framework. In Chile, the methodology of PISA assessment has also been used as a model for improving national assessment. Many countries have begun new teacher education projects, designed to promote mathematics education that better equips students for their futures in response to lessons learned from PISA. Some countries, including France and Denmark, have used the resources provided by PISA in these and other projects, especially using PISA items as a model for assessment items or a source of ideas for more complex items that share a PISA philosophy. Greater complexity and depth, and a fuller assessment of all phases of the modelling process is possible when items are to be used away from the very demanding context of the multi-country, multi-language, tightly-timed PISA survey. Some contributions, including those from Iran and Indonesia, also highlight classroom activities for students.

These contributions also show the impact of the PISA Mathematics Framework on thinking about the goals of mathematics education and the conceptualisation of the mathematics curriculum. A strong theme is the desire and need in many countries to give more emphasis to PISA's mathematical literacy with its emphasis on mathematics for all citizens across all parts of their lives. However, it is also the case that there has been considerable thought generated about the adequacy of mathematical literacy as a goal of mathematics education and how this can or should be balanced in a school mathematics curriculum with attention to intra-mathematical goals such as mathematical structure and attention to mathematics as a discipline studied for its own interest and beauty. Several contributions, including from Israel and Korea, report on the thinking stimulated by PISA ideas within curriculum review processes. For example, in Korea, a new series of textbooks gives more attention to contexts through a 'story-telling' approach that presents real or fantasy contexts to motivate and illustrate mathematical principles. This resonates with the 'educational modelling' approach outlined by Stacey in Chap. 3. Fundamental debate about the nature and goals of a good mathematics curriculum has also been a feature of the response in the USA.

An important aspect of the impact of PISA on thought about mathematics education has come through the prominence that PISA has given to mathematical competencies (called the fundamental mathematical capabilities in the PISA 2012 framework). Several contributions, including from Spain, report how these have been used to guide curriculum and assessment, and how the competency view, described more fully by Niss in Chap. 2 of the present volume, has been consistent with other influential initiatives in the early years of this century. The contributions from France, Indonesia and Chile also record the incorporation of PISA-like mathematical competencies in revised curriculum priorities.

In summary, these reports show that since its inception, PISA has had substantial influence on developments in mathematics education through the monitoring of performance, by the resources produced, and through the stimulus to fundamental reconsideration of the goals of mathematics education that is offered by the various components of the PISA mathematics framework.

Chile

About the Contributor

Felipe Almuna is currently a Ph.D. student in Mathematics Education at The University of Melbourne. After a career as a secondary and tertiary mathematics teacher in Chile he decided to undertake further studies in mathematics education. In 2010 he was awarded his master degree at The University of Melbourne, studying how the context influences students' approaches to PISA-like problems and winning the John and Elizabeth Robertson Prize for best research essay. In 2011 and 2012, he worked again as a teacher in Chile. His doctoral research is studying the relationship between contextualisation of mathematical problems and students' performance.

PISA: A Referent for Improvement

Chile has participated in four PISA survey administrations. Participation in the 2000 administration was delayed until 2001, and then the country participated normally in 2006, 2009, and 2012. The PISA 2009 survey ranked Chile in 49th place for mathematics among 65 participating countries and in the second place in the Latin American region after the partner country Uruguay. The mean score of 421 points is 75 points (three quarters of a standard deviation) below the OECD average of 496 points (OECD 2010b).

Aside from the rankings, the PISA 2009 mathematics results revealed that less than 1 % of Chilean students reach the highest level of proficiency in mathematics with scores higher than 669 points, and 51 % of students perform at or below the lowest level of proficiency with scores between 358 and 420 points (OECD 2010b). These results confirm that Chile still lags behind the OECD average and that there remains considerable action to be taken in matters related to education. At the time of writing, the first PISA 2012 results are available, showing an average score of 423, a small but statistically significant improvement.

Chile is taking steps designed to improve the quality of education; raising educational standards in Chile is "high on both the public and government agenda" (OECD 2010b, p. 87). In this way, the PISA results have been used as a referent to monitor variations of the educational goals in order to advocate policy change, promote educational research, and learn lessons from the PISA survey methodology.

In this vein, the national mathematics curriculum implemented in the 2000s has been reviewed. Since 2009 a greater emphasis on the notion of mathematical competency and mathematical reasoning (Solar et al. 2011) is observable in it. This curricular review in mathematics has taken into account revisions and analyses of curricula from OECD countries as well as frameworks and evidence from TIMSS and PISA (Ministerio de Educación 2009).

In addition, PISA has also started to influence educational research in Chile. In 2011 the Research and Development Office (FONIDE, standing for its Spanish acronym), a section of the Ministry of Education, launched a special round of grants for researching the impact of PISA in Chile and 25 % of the participating projects were related to PISA mathematics.

PISA assessment also has been influential in the improvement of the national assessment in Chile (SIMCE for its acronym in Spanish). PISA has been used as a best-practice guide to adapt existing assessments, in guiding methodological changes in SIMCE "improving procedure, manuals, item construction, statistical analysis and keeping records" (Breakspear 2012, p. 22).

Final Remarks

As Chile did not take part in the PISA 2003 survey (where the main focus was on mathematics) comparison in mathematics is only possible between the 2006 and 2009 survey administrations. The results show that since 2006 the results in mathematics did not change significantly. Hence, the influence of PISA mathematics in Chile has not yet been greatly evident. However, PISA mathematics has been a referent for the latest curriculum review in Chile. In 2009, the release of the PISA results in both reading and mathematics produced an immediate public concern. The analysis of the results of PISA has also been taken into account by policy makers when discussing the quality of the educational system in Chile. The PISA survey has offered to Chile an opportunity to raise critical questions about the learning outcomes, distribution of learning opportunities, skills and competencies that the Chilean educational system provides to students to equip them for today's globalised world.

Denmark

About the Contributor

Lena Lindenskov is from the Department of Education at Aarhus University in Denmark. She has worked in the Danish PISA Consortium since 1998 responsible for the mathematical literacy part. Lena also was the Danish representative in the PISA 2003 Mathematics Forum.

Alignment with Educational Goals of Denmark

From a Danish perspective, the Mathematics Framework from PISA 2000 has been of great interest, as it seems to be more applicable to the Danish mathematics education than the TIMSS survey. Mathematics in use, in everyday life, and for active citizenship is a priority for compulsory education in Denmark. The PISA definition of mathematical literacy and its further description seem to be in line with the intended goals and guidelines of Danish schools. Also the fundamental mathematical capabilities (OECD 2013a) underlying the mathematical processes resemble what is known in Denmark as the concept of the eight mathematical competencies, which are described by Niss (2003) and also in Chap. 2 of this volume. The concept was incorporated into national teaching guidelines from 2003 and into the national curriculum from 2009 as described in the Fælles Mål [English: Common Goals] (Ministry of Education 2003, 2009). The concept has been discussed in teacher training courses and applied in developmental projects and research around Denmark.

As the PISA Framework is in line with Danish educational goals, one might expect relatively high Danish performance. Throughout the PISA surveys 2000–2012, Danish students performed above the OECD average in mathematical literacy with means of 514, 514, 513, 503, 500 while the mean scores in reading literacy and science literacy did not differ from the OECD average.

Actions Following

The PISA performances of Danish students have had a big impact on educational and political debate and decisions, as noted by Egelund (2008). New critical questions were raised about the level of performance and about social-economic, ethnic and gender equity factors, considering that Denmark is a rich state with a strong emphasis on social welfare. Following an international OECD review in 2004 (Mortimore et al. 2004), national tests were introduced in several subjects, for example, in Grades 3 and 6 mathematics. For the first time the national teacher guidelines for mathematics in 2003 and 2009 included sections on students with special needs, influenced by several factors including the PISA results. They showed that although the number of low performers is small relative to international figures, in a national context the number is considered to be too high.

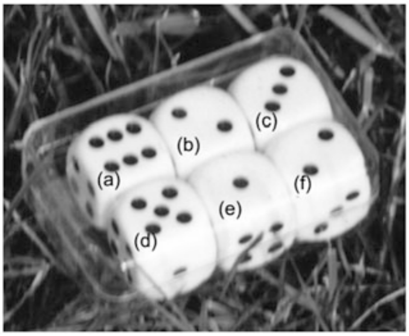
As PISA items are well described, the Danish mathematics team for PISA decided to investigate how released items can be used by teachers for formative assessment of their students and as ideas that they can develop into learning activities. We made secondary analyses of the student answers to released PISA-items based on single item statistics, together with an in-depth analysis of the written work of large samples of Danish students in PISA 2003. The results were published on the web with the title *15 Mathematics Tasks in PISA* (Lindenskov and Weng 2010). Our aim was to present rich descriptions and examples of how Danish students answer mathematics tasks when they participate in PISA surveys. We wanted to give descriptions that were rich enough for teachers to be able to relate them to their own practice. We looked into four PISA units in *Space and shape*, two units in *Change and relationship*, five units in *Uncertainty* and four units in *Quantity*. The items in these units covered all levels of difficulty.

The item *M145 Cubes*, as shown in Fig. 15.1, is categorised as *Space and shape*. The difficulty level is low, and student answers are coded in PISA just with one digit as full credit (in this case the correct answer of 1, 5, 4, 2, 6, 5 in order) or no credit for any other answer. (For further information about coding, see Chap. 9 by

Question 1: CUBES *M145Q01*

In this photograph you see six dice, labelled (a) to (f). For all dice there is a rule:

The total number of dots on two opposite faces of each die is always seven.



Write in each box the number of dots on the **bottom** face of the dice corresponding to the photograph.

Fig. 15.1 PISA released item Cubes M145Q01 (OECD 2006)

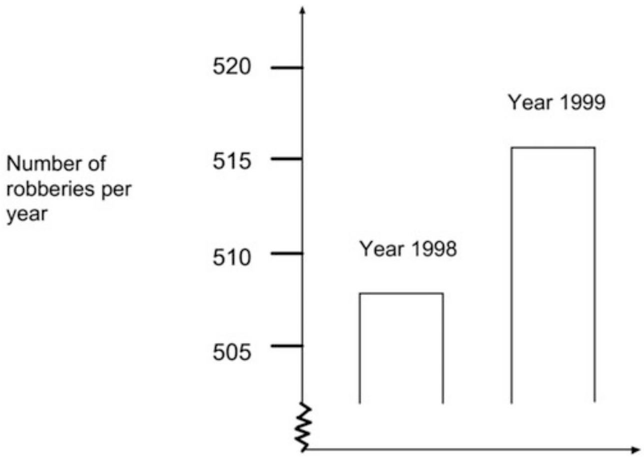
Sułowska in this volume.) We looked more deeply into 110 Danish student answers. We found three kinds of incorrect answers, and we created three second digit codes. Some students *copied* the numbers shown on the dice (answering 6, 2, 3, 5, 1, 2) another group *mirrored* them (answering 5, 1, 2, 6, 2, 3), while a third group made *calculation errors*. The two first types of answers indicate conceptual misunderstanding of 'the opposite side of a cube', despite the algebraic rule of the sum being given as a hint, and the third one indicates arithmetic problems. In assessment for learning we suggest the use of tasks like *M145 Cubes* in order to find indicators of students' thinking.

It is our general impression that for items with short answers, the most interesting information for teachers is the different types of incorrect answers. Concerning items with extended answers, it is also interesting for teachers to look into the different types of correct answers. The unit M179 Robberies (shown in Fig. 15.2) is classified in *Uncertainty*. The difficulty level is high. Full credit answers are coded in PISA with three double digit codes. Partial credit answers are coded with two double digit codes. No credit answers are coded with four

Question 1: ROBBERIES *M179Q01- 01 02 03 04 11 12 21 22 23 99*

A TV reporter showed this graph and said:

"The graph shows that there is a huge increase in the number of robberies from 1998 to 1999."



Do you consider the reporter's statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.

Fig. 15.2 PISA released item Robberies M179Q01 (OECD 2006) double digit codes. OECD (2006) gives full details of the coding criteria. All nine double digit codes are represented among the Danish student answers, which we looked into further. We saw that the full and partial credit answers were longer than the no credit answers. We saw that more everyday knowledge and less mathematical knowledge were used in the no credit answers than in the other answers. The diversity of the answers—in addition to being correct, partially correct or incorrect—shows the complexity of the item, and it seems that M179 Robberies motivates students to engage in interpretation and in reasoning. Here are some examples of answers given by students in Denmark, translated by the contributor.

- Some development has taken place. We see more robberies, but not in any strong sense. It has grown with approx. eight robberies (found from the graph), and that is not very much. The journalist has exaggerated, but when you look at the graph it looks bad, but the 'titles' are close to each other, that is why a growth of eight robberies looks very big.
- Such a small growth may be random, and next year you may have a marked decline in robberies. So I think the interpretation is unreasonable.
- I don't think nine robberies is a very big growth.
- What do you mean? It is reasonable, but how can I show it?
- Reasonable. I suppose so, but you cannot precisely see how many burglaries there were in 1998. It would have been better with a line diagram.
- It would have been easier if you had shown it on a circle diagram instead.
- Yes, there is an increase, so it is a fine interpretation, but she is not reasonable when she says it is a huge increase.
- No, because it is not a huge increase, but you know journalists can say anything.
- No, it looks huge in the illustration; you see the relative height of the two columns, but looking at the numbers only an increase of about 9.

In our view, secondary analyses of this kind can support development of mathematics education away from looking at mathematical tasks as something that should be finalised with one right answer as quickly as possible towards looking at mathematical tasks as initiators for problem posing, problem solving, reasoning and communication. We have observed an interest among teachers in the secondary analyses we made. We have observed students' interest as well. Some successful students were interested in looking at different student answers, including those from other countries, while weaker performing students said they were afraid that they would get confused.

Although the concept of mathematical literacy in PISA is regarded as in line with main ideas for mathematics education in the compulsory years of schooling in Denmark, critical questions are frequently raised in the debate on the value of mathematics in PISA. For example, there is debate on whether PISA measures give valid indications of the level and structure of 15 years olds' readiness for acting and reflecting on mathematics in use.

France

About the Contributors

Franck Salles and Jean-François Chesné work together at the DEPP in Paris. Franck Salles works both as a mathematics teacher in secondary school and as a research fellow at the office for students' assessment, DEPP, Ministry of Education. Franck has shared the position of National Program Manager of PISA for France, and is the French National Centre mathematics expert for PISA 2012. Jean-François Chesné joined the DEPP, Ministry of Education, working on assessment after a career as a secondary mathematics teacher and at the University Paris 12 where he was in charge of initial training for mathematics trainee teachers and professional development for in-service teachers. He heads the office of the evaluation of educational activities and experimentation. He conducts research on teaching practices and students' skills in mathematics in compulsory schooling. He was a member of a national jury for the recruitment of mathematics teachers and is a textbook author.

The Common Core of Knowledge and Skills and Complex Assessment

Unlike some other OECD countries, France did not experience a 'PISA Shock' after the first results of PISA from the year 2000. Nonetheless, PISA led to questioning of the adequacy of what is taught in French schools, especially in respect to how students use their knowledge in real-life situations. Thus, at an institutional level, PISA has had an influence in shifting the nature of knowledge towards a more applied and useful one.

In 2006, the law addressing the future of schooling in France amended the lower secondary curricula and established a *common core of knowledge and skills* (Legifrance 2006). This reform explicitly states that it was based both on recommendations of the European Union regarding '*key competences for lifelong learning*' (European Communities 2007) and on the PISA Framework. PISA's notion of mathematical literacy is underlying the *common core* as is clear from its definition:

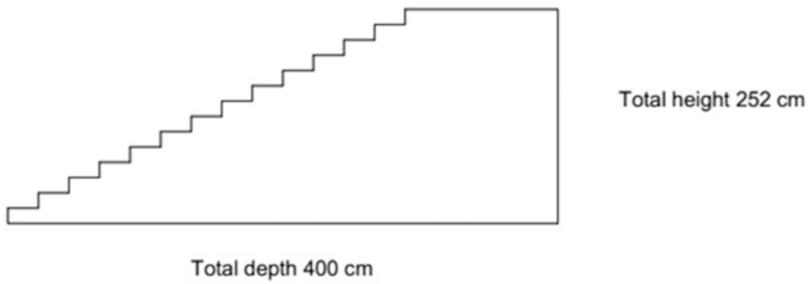
> knowledge and skills which are necessary to master at the end of compulsory education in order to successfully continue training, build one's personal and professional future and play a successful part in society. (Legifrance 2006, ANNEX)

As a result, skills and competencies connected with pure content have come to play a new and important part in curricula.

In mathematics, the *core* outlines skills such as *reasoning*, *communicating*, *implementing*, *handling information*, which are employed in four clusters of content (*numbers and operations*, *geometry*, *measurement*, *data handling/uncertainty*). This is very similar to the fundamental mathematical capabilities and the content

Question 1: STAIRCASE M547Q01

The diagram below illustrates a staircase with 14 steps and a total height of 252 cm:



What is the height of each of the 14 steps?

Fig. 15.3 PISA released item M547 Staircase (OECD 2006)

categories of the PISA 2012 Mathematics Framework (OECD 2013a) and its predecessors.

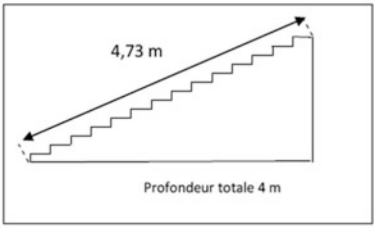
From 2008, new official instructions for mathematics teachers require developing and assessing students' skills within complex tasks through various contexts (MENJVA/DGESCO 2011). In addition to examining students' final productions, teachers must pay specific attention to their intermediate processes, partial reasoning, and spoken or written communication. This emphasis on reasoning and communicating is not only in geometry, as it often used to be, but also in arithmetic and algebra (MENJVA/DGESCO 2009a). Documents published by the Ministry of Education (see for example, MENJVA/DGERSCO 2009a, b) are often based on PISA released items. Figure 15.3 displays a PISA item M547 *Staircase* which was released after the PISA 2003 main survey (OECD 2006). The difficulty of the item is at Level 2, just above the boundary of Level 1. Figure 15.4 shows an adaptation (MENJVA/DGESCO 2011), illustrating the possibility of proposing a classical geometry problem in a real-life context. The French instructions translate as:

> For a staircase to conform to regulations, the height of each step must be between 17 cm and 20 cm. Does the staircase shown in the diagram meet these regulations? Show all of your working, even those paths which were not successful.

In the adaptation, the mathematical task has been made considerably more complex than the quite simple original, which involved only dividing the total height by the number of steps and ignoring the redundant information of 400 cm depth. In the new item, the given data was modified, Pythagoras's theorem is likely to be used, metres are to be converted into centimetres, and the question requires that the final value is tested to see if it fits in the specified range. As with many PISA

Exemple en mathématiques : l'escalier

Pour qu'un escalier soit conforme aux normes, la hauteur de chaque marche doit être comprise entre 17 cm et 20 cm. L'escalier représenté sur le schéma ci-contre est-il conforme aux normes ?



Tu présenteras ta démarche en faisant figurer toutes les pistes de recherche même si elles n'ont pas abouti.

Direction générale de l'enseignement scolaire
© Ministère de l'éducation nationale, de la jeunesse et de la vie associative

Page 4 sur 15

Fig. 15.4 Adapted Staircase Item (MENJVA/DGESCO 2011, p. 4)

items, an alternative solution method is also possible, in this case involving scale drawing, and this makes the item accessible to more students. These modifications make it a complex task meeting official standards.

One cannot claim that these directions have had wide and direct influence on actual teaching practices in France. This very innovative reform was not followed by widespread national teacher training. The evolution of teaching practices is a slow and complex process in the centralised French educational system and still today, most teachers are not familiar with PISA. However, intermediate institutions have been strongly influenced. Teacher trainers often mention PISA, its Framework, items and their coding guidelines during initial courses about the *core*. Textbook editors update mathematics textbooks to include more and more PISA-like *common core* situations. And last but not least, national inspectors are gradually modifying national examinations to include more complex tasks in context, and are valuing partial reasoning and different forms of communication.

Indonesia

About the Contributor

Professor Zulkardi is a lecturer in the Department of Mathematics Education in the Faculty of Teacher Training and Education, Sriwijaya University, South Sumatra, Indonesia. In 2002 he got his PhD on realistic mathematics education from the Netherlands. One of his supervisors was Professor Dr Jan de Lange, the first Chair of the PISA Mathematics Expert Group. Since then, Zulkardi has been involved in many projects related to PISA, some of which are discussed in this contribution. Since 2008, he has been the Vice President of the Indonesian Mathematical Society for Education and in this capacity he started the first international journal on mathematics education in Indonesia called IndoMS-JME (jims-b.org).

General Influence of PISA Mathematics in Indonesia

As do many governments that participate in PISA, the Indonesia government uses PISA to monitor the performance of the educational system. The purpose of this contribution is to present information and describe the ways in which PISA mathematics has influenced the thought and action of some groups of people in Indonesia. These groups are the central government, teacher educators and the PMRI team (Realistic Mathematics Education, Indonesia).

Since the PISA survey was first launched by the OECD in 2000, Indonesia has participated but its results, especially in mathematics, have been low, with some instability. First, in 2000, Indonesia was ranked 39 of 41 countries in mathematics. Then in 2003, the rank was 38 of 40 countries and in 2006, 50 of 57 countries. In 2009 it decreased to 61 of 65, and to 64 from 65 in PISA 2012 (although the mean score was the same).

Figure 15.5 shows the mean scores for mathematics, science and reading for Indonesia for the first four PISA assessments. One can see that there has been a steady increase in mean scores for the reading scale since 2000. The 2009 mean for science shows a drop of 10 points from a fairly stable level in the previous three

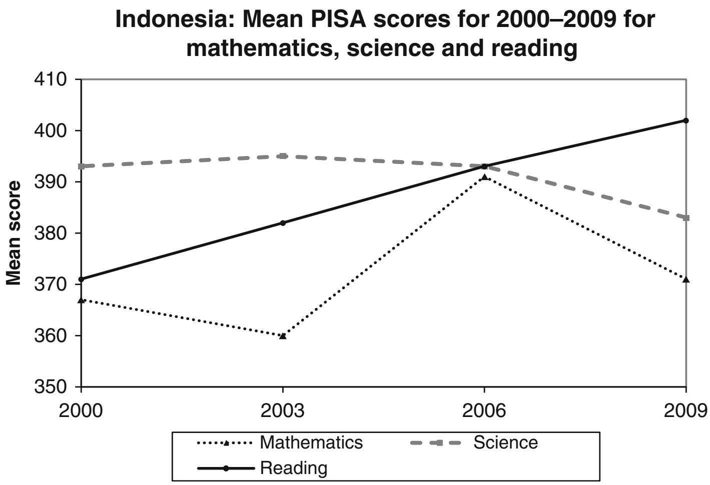


Fig. 15.5 Indonesia's mean PISA scores for 2000–2009 for mathematics, science and reading literacy (Stacey 2011)

assessments. The mathematics score has been more unstable. A different way of interpreting the data is that it has been steady, except for a relatively high score in 2006 (Stacey 2011).

PISA and RME in Indonesia

In July 2000, Professor Jan de Lange from the Freudenthal Institute in the Netherlands was invited as a keynote speaker in the National Conference on Mathematics at the Institute of Technology in Bandung. He presented new issues on mathematics education in the world, including PISA and Realistic Mathematics Education (RME). He also explained that the goals of mathematics education had changed from its earlier focus on mastering basic skills of mathematics with few applications. The new goals of mathematics education were to help students become good problem solvers and smart citizens.

A year later, the Freudenthal Institute and the National Centre for School Improvement (APS) both from the Netherlands, helped a group of Indonesian mathematicians and teacher educators headed by Professor R. K. Sembiring to set about reforming mathematics education in Indonesia. They adapted the Dutch instructional theory of Realistic Mathematics Education (RME) to its Indonesian version called PMRI (Pendidikan Matematika Realistik Indonesia). The PMRI project formally started in 2001 in four teacher education institutions and 12 primary schools in Java. By 2013, PMRI has been disseminated to the 23 of 33 provinces in Indonesia. More information about the project of PMRI can be seen at the PMRI portal http://p4mri.net and in the article by Sembiring et al. (2010).

In a 2007 national seminar on mathematics education in Palembang, Professor Fasli Jalal, the Director General of Higher Education presented, on behalf of the Minister of National Education of Indonesia, the PISA results for 2003 and 2006 on mathematics education. He urged the participants of the conference who were mostly school mathematics teachers, to learn from PISA results by improving the instructional quality and using PISA problems that had been released and were available on the web (OECD 2006, 2013b). Although that was only a suggestion, some people, including the contributor, were inspired to infuse the PISA spirit and use PISA problems in assessment and for research projects.

Zulkardi (2010) stated that there is a gap between the content of curriculum in Indonesia and the problems that were tested in the PISA mathematics. He also analysed the mathematics problems in the National Examination (UN). He found some mathematics problems were different to PISA. Most of the problems in the UN were in the low and middle difficulty levels of PISA. Therefore, he suggested to the government that some PISA-type problems should be included in the next UN so that students and teachers will be aware of the problems and these will automatically guide students to learn how to do PISA problems.

The Indonesian government has also used PISA results as one of several arguments for changing the mathematics curriculum to the new Curriculum 2013.

The PISA mathematics scores in 2009 show that the vast majority of Indonesian students are only able to understand mathematics up to level 3 of PISA, while significant proportions of students in many other countries reach levels 5 and 6. Therefore, it is assumed that the materials and the process of learning in Indonesia differ from those in developed OECD countries. Using PISA results as one of the arguments, the government of Indonesia changed the curriculum and the new curriculum was implemented starting from July 2013 at Years 1, 4, 7 and 10. The curriculum aims to include more problem solving, modelling and reasoning for mathematics and to use more information and communications technology for content and teaching delivery.

PISA for Students and Teachers

Kontes Literasi Matematika (KLM) is a national contest of mathematical literacy for high school students that began in 2010 (Widjaja 2011). The first KLM was initiated by the present contributor, Zulkardi, at Sriwijaya University working with about 200 junior high school students. The KLM contest begins by participants solving PISA-type problems in a written test, which is graded by a committee. Then, about 20 % of participants are chosen to compete in the semi-final, where participants have to explain their solutions or strategies in solving the problem. Lastly, from three finalists, the champion of mathematics in the province is selected.

In 2011, the second KLM was conducted in seven big cities namely Medan, Palembang, Jakarta, Yogyakarta, Surabaya, Banjarmasin and Makassar. In 2012, the contest added five new cities: Padang, Semarang, Malang, Kupang and Ambon. For the last 2 years, the grand championship of KLM has been conducted at the National Training Centre of Mathematics Education in Yogyakarta. The winners from each city participate in this national competition.

PISA results have slowly influenced the curriculum of mathematics education in teacher education. For instance, PISA has been part of the content in an assessment course at the Department of Mathematics Education Graduate Program at Sriwijaya University in Palembang. In this course student teachers learn what PISA problems are and how to design PISA problems using real-life contexts from Indonesia. Based on that course, some student teachers are doing research projects about how to design PISA-like problems.

Information About PISA

PISA was seen as newsworthy as soon as the national scores were released. For instance, Kompas, the biggest newspaper in Indonesia, has always published the PISA ranking, along with expert commentary on the PISA results and their implications for future leaders of Indonesia. Two sample articles are "70 % of Indonesian students will find it difficult to live in the twenty first century" (Erlangga 2012) and "Why Indonesian students have low achievement" (Nurfuadah 2013). However, little action followed their comments.

PISA mathematics in Indonesia has also featured in IndoMS-JME (http://jims-b.org) the *Indonesian Mathematical Society Journal on Mathematics Education*. One good article is an invited article written by Kaye Stacey (2011). Several other IndoMS-JME articles about PISA problems have been contributed by Zulkardi's research students (i.e. Kamaliyah et al. 2013). There is also a supplementary book (Wardhani and Rumiati 2011) on instrument evaluation for mathematics achievement that draws on both PISA and TIMSS, which has been prepared in the context of the project BERMUTU. In addition to the journal and news, the contributor has also designed a blog (http://pisaindonesia.wordpress.com/) that provides information about PISA Indonesia, PISA released problems, PISA-type problems and links to other blogs relating to PISA.

Summary

In summary, thinking about mathematics education has been substantially influenced in Indonesia by the ideas championed by the Freudenthal Institute and elsewhere about the need for realistic mathematics education. These ideas have been well publicised and made concrete by the PISA tests. Indonesia's poor results provide a challenge to the nation, which is being addressed in part by using PISA items as models for teaching and learning.

Iran

About the Contributors

Professor Zahra Gooya and Dr. Abolfazl Rafiepour are active contributors to mathematics education in Iran. Zahra Gooya from Shahid Beheshti University is the first mathematics educator to have had an in-depth influence on mathematics education in Iran. A celebration of her 20 years of contribution was recently organised by her colleagues. She has often written about international studies in the national journal, and many teachers have become familiar with these international developments through this path. Dr. Abolfazl Rafiepour, previously a secondary school mathematics teacher, was one of the first students to start a master of mathematics education under Professor Gooya in 2001. His master and doctoral theses analysed TIMSS data. In addition to his other work at Shahid Bahonar

University of Kerman, he is now director of Kerman Mathematics House, the second one to be established in Iran.

The Influence of PISA in Iran

Iran has participated in TIMSS since 1995, but not in PISA. Even though it has not participated as a country in PISA surveys, the PISA study has had a considerable influence on mathematics education research in Iran. This contribution documents some of the actions and changing thought that is evident in the work of teachers, mathematics education researchers, student activities and textbooks.

A number of master degree research studies from primary to tertiary levels have concentrated on mathematical modelling and applications, which is one of the focal points of PISA. Some papers are in Persian (or Farsi, the official language in Iran) including Ahmadi and Rafiepour (2013), Faramarzpour and Rafiepour (2013) and Karimianzadeh and Rafiepour (2012). There are also some papers in English that focus on modelling and applications from the Iranian students point of view, such as Rafiepour et al. (2012), Rafiepour and Abdolahpour (2013) and Rafiepour and Stacey (2009). There have also been presentations at the annual Iranian Conference on Mathematics Education, including in 2012 papers by Abdolahpour, Rafiepour and Fadaie on the level of mathematical modelling competence of students and by Esmaili, Esmaili and Rafiepour on the effect of different types of problems on students' emotions.

In addition, many interested graduate students have produced papers based on modelling activities that they have conducted with school children and have presented them at mathematics education conferences in Iran. Almost all these graduate students are mathematics teachers and they work voluntarily with students providing extra-curricular activities in the Mathematics Houses across Iran. Their main purpose is to bridge the gap between school and real-life mathematics and to promote mathematical literacy.

Since 2004, the first 10 days of the eighth month of the Iranian (Jalali) calendar (22–31 October) have been named the "Mathematics Decade" by the Iranian Mathematics Society. During this time, all Mathematics Houses are actively involved in out-of-school activities to promote mathematical literacy. Many students, teachers and ordinary people visit the Mathematics Houses and other related organisations and get involved with mathematical activities. To give an example, in 2011 and 2012, the Kerman Mathematics House used some of the PISA released items (OECD 2006, 2013b) related to modelling and applications during Mathematics Decade. Students were actively engaged in doing mathematics and enjoying it. The main purpose of these modelling activities was preparing students for using their mathematical knowledge together with their daily experiences to solve real-world problems.

Another effect of PISA is that policy makers claim that it has influenced the direction of change in the new national mathematics textbooks. However, the reality of this claim has been questioned by Gooya (2013) and Hasanpour and Gooya (2013). Their view is that mathematical literacy and real-life activities are not promoted only by the inclusion of real objects and phenomena in textbooks, but "realistic mathematics education" situations must be created where students are involved in solving problems in genuine real-world contexts. This will include some modelling activities. The present contributors have examined the way in which the new mathematics textbooks for Grade 9 students might cultivate mathematical literacy (Rafiepour et al. 2012).

To sum up, school mathematics in Iran has been implicitly influenced by the PISA rationale via different genuine activities that are designed and carried out by some mathematics teachers and educators. Presenting this new direction for mathematics education has created new opportunities for young researchers as well as bringing some hope for the former generation to think more seriously about the feasibility of what Freudenthal preached a long time ago about 'Realistic Mathematics Education'. At the formal policy level, despite the claims, nothing much has yet been done to address the deeper issues of mathematical literacy.

Israel

About the Contributor

Dr. Hannah Perl works for the Ministry of Education in Israel. She served for many years as the highly-respected Chief Inspector for Mathematics in the Ministry of Education, where all major decisions about mathematics, including curriculum, testing, and teachers, were her responsibility. She is now the head of the science division in the pedagogical secretariat of the Ministry, which includes supervision of all science and mathematics education. She has undertaken various research projects including very interesting research with graphing calculators long before the use of technological tools was in the headlines.

The Influence of PISA on Mathematics Education in Israel

In Israel, mathematics has always been an obligatory part of the school curricula beginning in kindergarten and continuing throughout the 12 grades of the school system. One of the traditional arguments in support of this decision (among other important ones) has been that mathematics, because of its abstractness and special reasoning tools, is a universal means for describing the world around us and thus constitutes a necessary ingredient of every student's problem solving tools. It was believed that equipping students with these tools suffices to ensure that they would be able use them whenever necessary to solve problems in a variety of contexts.

Although the middle school and high school curriculum stated the importance of developing students' ability to decide when and how to use mathematical concepts, actual teaching practices in schools emphasised traditional mathematical skills and understanding and did not implement the developing of students' ability to apply their mathematical knowledge to solve authentic problems in a wide range of situations.

The results of international surveys and assessments such as TIMSS and PISA have underscored the fact that the ability to identify and apply mathematics when it is needed does not develop by itself, even in mathematically oriented students, and has to be taught explicitly to both mathematically strong students and those who are not mathematically inclined. Thus mathematics education policy makers and curriculum developers in Israel were challenged to re-examine the mathematics curricula (Grades 7–12) and to rethink it in terms of the content, skills, processes and contexts that have the potential to bring our students to achieve mathematical literacy as defined by PISA.

There was a debate regarding the role of mathematical literacy in teaching mathematics to all students. It became necessary to answer the questions "What mathematics should be taught?", "To whom?" and "How?" The utilitarian approach was important but not acceptable as the main or only organising theme of the curriculum. Other traditional considerations that were considered equally important were teaching mathematics for intellectual pleasure, noticing the aesthetics of mathematics and appreciating it as an important cultural achievement of mankind, understanding abstract structures, solving pure mathematical problems, and developing high order thinking skills. The Mathematics Professional Advising Committee to the Ministry of Education revised the middle school curriculum taking all these aspects into consideration.

In middle schools (Grades 7–9) mathematical literacy has become a part of the new curriculum for all students. Curriculum developers and textbook writers have broadened their traditional approach to school mathematics and realised that it is possible to find meaningful, interesting and authentic applications that are mathematically challenging for different grade levels and students' capabilities. Formal mathematics competency was not abandoned but reduced in size and relegated to the higher grade levels. It was also understood that in order for students to effectively deal with these new tasks, teaching practices must change and learners will have to be taught in new ways that, hopefully, will raise the learning and teaching standards and also support intellectual enjoyment for all. Resources were made available to implement these changes. They included the design of new teaching and learning materials, teacher professional development and the appointment of school instructors to assist teachers in the classrooms.

In high school (Grades 9–12) a new mathematics curriculum is currently under development. Mathematical literacy will be taught to all students but in different ways at different levels depending on students' mathematical abilities and inclinations. Students who are not mathematically inclined will focus on mathematical literacy with higher mathematics content so that they will be able to autonomously engage a wide range of real-life mathematical and basic statistical situations. For mathematically oriented students the concept of mathematical literacy will be broadened to include not only real-life situations but tasks that are more complex and abstract and which integrate a larger range of topics (including applications to other scientific disciplines), the reading of advanced mathematical texts and use of higher level mathematics concepts and competencies. Levels of performance will be in accordance to the six levels defined in the proficiency scale descriptions of the PISA Framework.

All mathematics curricula will incorporate use of twenty first century technology both in learning and assessment. Details of the curriculum changes in the middle school can be found (in Hebrew) on Israel's Ministry of Education website: http://cms.education.gov.il/EducationCMS/Units/Mazkirut_Pedagogit/Matematika/ChativatBeinayim/.

Korea

About the Contributor

Kyungmee Park is a professor at Hongik University in Korea, teaching pre-service teachers. She was a member of the PISA Mathematics Expert Group from 1998 to 2004, and worked as a researcher at the Korean Institute of Curriculum and Evaluation, responsible for PISA 2000 in Korea. She is involved in mathematics curriculum and textbook development, writes mathematical columns in several daily newspapers, and has contributed to the popularisation of mathematics for the general public.

Impact on Mathematics Curriculum

The impact of PISA on mathematics education in Korea can be discussed in the two aspects of curriculum and textbooks. The Korean Institute of Curriculum and Evaluation (KICE), which is responsible for the development of mathematics curriculum in Korea, was heavily influenced by OECD's DeSeCo project (Rychen and Salganik 2003). DeSeCo is an abbreviation of 'Definition and Selection of Key Competencies'. Over 3 years, KICE attempted to similarly identify key competencies for Koreans of the future (KICE 2009). As a result, ten core competencies were identified: creativity, problem solving, communication skills, information processing, interpersonal relations, self-management, basic learning skills (literacy), citizenship, global awareness and vocational development. These competencies suggested directions for constructing national curriculum. However, the new mathematics curriculum of Korea announced in 2011 did not explicitly mention these key competencies. Instead, it emphasised the processes of doing mathematics. The mathematics curriculum states:

> Crucial capabilities required for members of a complex, specialised, and pluralistic future society are believed to be fostered by learning and practising mathematical processes, including mathematical problem solving, communication, and reasoning. (Ministry of Education, Science, and Technology 2011, p. 2)

In fact, problem solving, communication, and reasoning had already been mentioned in the previous mathematics curriculum, but the 2011 curriculum put more emphasis on them and intends to implement these three mathematical processes in the content. This emphasis can be interpreted as an influence of OECD DeSeCo and PISA. In particular, the mathematical processes are part of the mathematical competencies presented in the PISA 2009 Mathematics Framework (OECD 2010a).

Impact on Textbooks

The 2011 national mathematics curriculum emphasises contextual learning from which students can grasp mathematical concepts and make connections with their everyday lives. Thus the new textbooks developed for this curriculum include more real-life contexts. In addition, a 'story-telling textbook' was introduced as a prototype for mathematics textbooks. Story-telling mathematics textbooks have already been developed and are being used in Grades 1 and 2 from 2013. In the middle school and high school, the story-telling approach has been recommended to be adopted for textbooks and sample chapters have been prepared.

Here is an example. The chapter on "Measuring Length" in Grade 2 is called "The emperor's new clothes" (MEST 2013). The plot for the story is to make clothes for the King to wear on his birthday. Students play the role of the king and tailors, and they come to see the necessity of having standard units for measurement because otherwise the measurements vary from one tailor to another. Students naturally acquire the concept of standardised units through problem solving in this fairy tale. By learning mathematics through story-telling textbooks, students are expected to understand a concept in conjunction with a story that provides a practical impetus for and application of the concept. In the meantime, mathematical processes such as problem solving, communication, and reasoning are naturally embedded in each chapter (Kwon 2013).

Figure 15.6 shows three pages from the chapter "Measuring Length". On page 134, two tailors measure the length of the arms of King by using their palms. The male tailor on the left says "two palms" and female tailor says "three palms". Here, students are expected to think about the problems caused by these arbitrary body units to measure length. On page 137, students measure objects in the classroom using various body units. Before the metric system, body units such as feet were prevalent. Through this activity, students indirectly experience the historical development of measuring units. On page 150, the king and the tailors agree to introduce

Fig. 15.6 Sample pages from story-telling textbook (MEST 2013) (Reproduced with permission)

the centimetre to measure length as a standard unit. Students are expected to appreciate this uniform unit, which can be used in any place without confusion.

The PISA assessment takes a broad approach to measuring knowledge, skills and attitudes, moving beyond the school-based approach towards the use of knowledge in everyday tasks and challenges. Thus, despite often using fantasy settings, the story-telling textbooks are putting into practice the context-oriented nature of the OECD PISA philosophy.

Singapore

About the Contributor

Professor Berinderjeet Kaur is a professor of mathematics education and Head of the Centre for International Comparative Studies at the National Institute of Education in Singapore. Since 1995, she has been involved in the secondary analysis of TIMSS data for Singapore and other countries. She was the mathematics consultant to TIMSS 2011 and is presently a member of the Mathematics Expert Group for PISA 2015.

Affirmation of Mastery and Directions

Singapore participates in international studies to benchmark itself internationally and to learn from best practices of other education systems. Singapore has participated in TIMSS since 1995 for both Grades 4 and 8. The results of every administration of TIMSS for Singapore have affirmed that students have mastery of content knowledge according to international standards. In addition they are highly proficient in the application of their knowledge and in reasoning with their mathematics.

Although the first administration of PISA was in 2000, Singapore did not participate in PISA until 2009. As Singapore is a small country with only about 170 secondary schools, support must be obtained from all the schools as such international benchmarking studies require the participation of at least 150 schools.

The results of PISA 2009 Mathematics showed that Singapore was ranked second to Shanghai. The positive outcome affirmed that 15-year-olds in Singapore were able to apply reason and transfer their knowledge of mathematics in new, unfamiliar contexts, and demonstrate the ability to think critically and solve real-life problems. This outcome has affirmed that the systemic adoption of the "Thinking Schools, Learning Nation" vision (Goh 1997) for all schools in Singapore has had the desired and valued impact where students are acquiring the knowledge and skills necessary for the workplace.

Irrespective of the results in TIMSS and PISA, the mathematics school curriculum is revised every 6 years. The revision is guided by global developments, the needs of and feedback from stakeholders (including teachers and school leaders), as well as developments in the teaching, learning and assessment of mathematics. This allows the curriculum and resulting classroom practices and assessment modes to be revised periodically so that they remain relevant for students and for the economy.

Spain

About the Contributors

Luis Rico, José Luis Lupiáñez and Rosa M. Caraballo all work at the University of Granada in Spain. Dr Rico has been Professor of Didactics of Mathematics at the University since 1992, where he leads the Research Group on Didactics of Mathematics. He was member of the Mathematics Expert Group for PISA 2003. His main subjects of research are the design and development of mathematics curriculum, quality of mathematics training programs and quality indicators for mathematics education. In 2012 he was awarded the Social Sciences Research Prize "Ibn-Al-Khatib", by the Government of Andalusia. Dr Jose Luis Lupiáñez is a lecturer at the Mathematics Education Department of the University of Granada (Spain) where he teaches prospective primary and secondary teachers. His research focuses on teachers' learning processes, mathematics teacher training, mathematical competences and learning expectations. Rosa M. Caraballo, a Puerto Rican research student at the University of Granada, completed a master's dissertation on Spanish National Assessment tests in 2010, which are based on the PISA Mathematics Framework. Her doctoral dissertation is on mathematical tasks to assess mathematical literacy.

Mathematical Competency and the Spanish Curriculum

In 2006 the Spanish Education Fundamental Act (LOE, its Spanish acronym) was first passed and it remains in force. The Act proposed an evolution of the educational orientation in Spain and improvements to be followed in the succeeding years. The LOE responds, first, to the social changes of recent decades and to the demands of Spanish citizens for a general and democratic education. Second, it attends to the trend towards high quality education, which is acclaimed by the countries of the European Union in their agreements since the late twentieth century (Ministerio de Educación y Ciencia 2006).

As a definite and innovative tool, the Act introduced the concept of competency at all educational levels in the curriculum taking an inherently wide general conception. The Act defines curriculum as "the set of objectives, key competencies, pedagogic methods and assessment criteria outlined for each one of the subject areas the law regulates" (Ministerio de Educación y Ciencia 2006, p. 17166).

The LOE was grounded on the concept of lifelong learning. Education is perceived as an ongoing and dynamic learning process of progressive qualification.

> Everyone should have the opportunity to learn throughout life, in and out of the educational system in order to acquire, update, add to and expand his or her competencies, knowledge, abilities, aptitudes and skills for personal and professional development. (p. 17166)

Following the LOE provisions, the education system aims to provide students with the knowledge and skills necessary to perform effectively in the society of which they are part, in mathematics as well as in other subjects. Key competencies set these expectations for learning and training based on the DeSeCo (OECD 2005) and the Eurydice Projects (Unidad Europea de Eurydice 2002).

The Spanish curriculum does not use *mathematical literacy*; instead it uses the (parallel) term *mathematical competency*. The reason for this change of name is discussed in Chap. 1 of the present volume. Mathematical competency is considered to be one of the main basic learning expectations of the whole Spanish educational system. It should be understood as similar to mathematical literacy as defined by PISA Mathematics Frameworks for 2003 and 2012 (OECD 2003, 2013a), and the associated ideas of Niss (2002).

Diagnostic Assessments

On lifelong learning and basic competencies development, the LOE stipulates that diagnostic assessments of key competencies will be carried out at the end of the fourth course of primary education and the second year of compulsory secondary education (Ministerio de Educación y Ciencia 2006). They are preliminary and complementary to the PISA assessment; it is expected they will provide useful information to establish the progress of key competencies, especially the mathematics one as the law regulates.

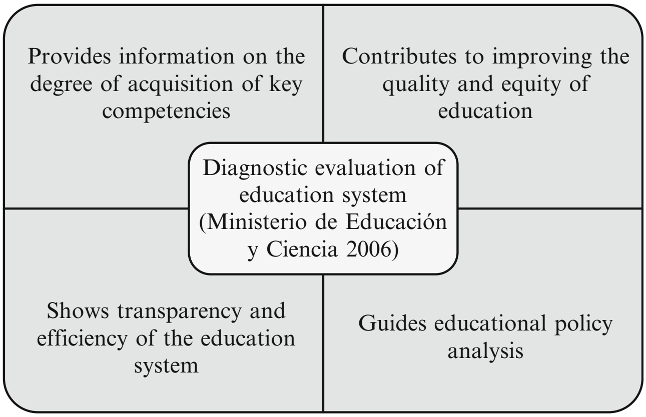


Fig. 15.7 Main goals of diagnostic evaluations of the Spanish education system

It is important to stress that the objective of these assessments is not to determine whether, and to what degree, the intended curriculum has been implemented. Rather, it aims to know the students' ability to apply their acquired learning when facing tasks that require them to cope with real-life situations. In addition, changes in the curriculum and key competencies introduced by the LOE, allocate priority to learning expectations. Figure 15.7 summarises the main goals of the general assessments.

Competencies and Mathematical Literacy Assessment

For mathematics in particular, diagnostic evaluations serve as training for the mathematical literacy evaluation that will take place at the end of the compulsory period through PISA. Here we can establish links between mathematical competency development and mathematical literacy at the end of compulsory education.

In order to assess mathematics competency, diagnostic assessments consider three dimensions: (1) the situations and contexts in which the competency is applied, (2) the processes that enable the student to apply the acquired knowledge to the contexts, and (3) the curricular content embedded in the full range of students' knowledge and skills. Of these three dimensions, the description of the contexts and processes are shared with the PISA Framework, whilst the content is described in terms of traditional curriculum areas rather than the *overarching ideas* of the PISA 2003 Framework (and the content categories of PISA 2012 Framework).

The link between PISA assessments and quality indicators for the Spanish education system is based on the notion of competency as a central concept (Rico 2011). There is a quality indicator (R2.2) for the second year of secondary school that is measured by the overall results achieved in the mathematical competency in the general diagnostic evaluation described above. The indicator for age 15 in mathematics (R3.2) is determined by the results of the international PISA study. Because it is included in the Education Quality Indicators, together with the national and regional diagnostic tests (Instituto de Evaluación 2011), the PISA assessment is very important in the Spanish educational system.

PISA Results

Spain has participated in all five PISA assessments that have been conducted so far. Table 15.1 presents the number of participating Spanish students and their average score in the four PISA assessments from 2000 to 2009, in the three main key competencies. The OECD average score was initially set at 500 with a standard deviation of 100. All of the average scores for Spain are below the OECD average, including the score (484) for PISA 2012. With a standard deviation of 100, approximately two thirds of students across the OECD score between 400 and 600. The number of students tested has been increasing in successive PISA administrations because of a desire to obtain reliable estimates of the performance of regional communities within Spain.

The poor performance of Spanish students in recent international comparative assessments, including PISA, has created widespread public concern. As a response, deep curriculum reforms were requested. In recent years, the results have systematically generated a major media debate that has often placed political blame on the incumbent government and emphasised the more negative aspects (Aunión 2007; Díaz and Suárez 2010). Notwithstanding, critical analysis that highlights achievements in addition to detecting deficiencies has been also carried out. Moreover, outcomes have been analysed from a constructive point of view (Recio 2010). As stressed by Rico:

> You have to understand and explain why Spanish results in PISA assessments are not satisfactory and therefore, channel the discussion towards the adoption of radical, urgent and appropriate measures to improve the curriculum and teacher training in mathematics. (Rico 2011 p. 10)

Recently, the Spanish Federation of Teachers of Mathematics organised a meeting aimed to study the design, organisation and impact of national and

Table 15.1 Number of participating Spanish students and their average scores in PISA assessments

Year	Average score			
	Number of students	Reading	Mathematics	Science
2000	6,214	493	476	491
2003	18,000	481	485	487
2006	20,000	461	480	488
2009	26,000	481	483	488

international assessments in Spanish mathematics education. They found that poor coordination of the various professional and government sectors involved in this process have great impact on the teaching and learning of mathematics.

Final Remarks

The impact of PISA has affected the foundation and organisation of the compulsory mathematics curriculum in Spain. The results of the evaluations raise questions about the quality of the system and show weak approaches to incorporating core competencies in school practice. Social concern is evident and the interest of parents and teachers to adopt corrective measures is strong. As in other countries, there has been no questioning of the learning model established by PISA.

There are favourable conditions for improving the institutional assessment system, involving both the general public and professional sectors. We must remember that PISA does not evaluate students or teachers; PISA provides indicators on the quality of the system. Everything is ready to improve the level of Spanish mathematics education.

United States of America

About the Contributor

Solomon 'Sol' Garfunkel is an American mathematician who has dedicated his career to mathematics education. Since 1980, he has been the executive director of the Consortium for Mathematics and Its Applications (COMAP), an award winning non-profit organisation that creates learning environments where mathematics is used to investigate and model real issues in our world. One acclaimed product is "For All Practical Purposes: An Introduction to Contemporary Mathematics", a television series and now textbook. Dr Garfunkel was a member of the PISA 2012 Mathematics Expert Group. In 2009, he was awarded the Glenn Gilbert National Leadership Award from the National Council of Supervisors of Mathematics.

An American Reminisces on PISA

First, to put this reminiscence in context, I should state that I was a 'math warrior', from what I regard as the losing side of the 'math wars' that raged in the United States especially during the 1990s and continue to some extent today. For readers unfamiliar with these issues, Schoenfeld (2004) provides a history of the debate and

Harwell et al. (2009) is one reference discussing the hotly contested differences over approaches to mathematics curriculum and teaching.

My background in mathematics education is in curriculum reform. I have been involved in the creation of literally hundreds of modules, textbooks, and one comprehensive 4-year secondary school curriculum. All of these exemplify the importance and centrality of mathematical applications and modelling. They are about teaching mathematics through its contemporary use. And they are in the spirit of the 1989 NCTM standards. Without rehashing the issues of the 'math wars', it is fair to say that the approach of the 1989 NCTM standards has now been supplanted in the U.S.A. by the new Common Core State Standards in Mathematics (CCSSM 2010). While applications and modelling get a nod in these standards, they are certainly not as central as arithmetic and algebraic fluency and the exposition of mathematical structure. I have been an outspoken critic of the CCSSM, although I am working with a number of organisations to make standards implementation go as smoothly as possible—for our students' sake. One such group is *Achieve* (www.achieve.org), a non-profit organisation set up to provide technical assistance and research capacity to U.S. states on educational reform, especially standards, assessments, curriculum and accountability systems. I have consulted for *Achieve* on a number of projects. I am usually seen to be on the philosophical 'left', balancing off other consultants who occupy space on the philosophical 'right'.

Now, I have kept up with PISA and the work of the Mathematics Expert Group (MEG) through personal friends and colleagues since 2003. As a consequence I was aware that PISA had come in for some criticism from some members of the mathematics research community for not being 'mathematical' enough. This criticism by and large came from conservative 'math warriors', and clearly the OECD's PISA Secretariat was sensitive to their comments. *Achieve* was brought in to assist the international contractors with the preparation of the Framework for mathematical literacy for 2012, as well as conducting an international consultation on the earlier and proposed frameworks and external validation of the alignment of the final item pool to the agreed framework and the presence of explicit mathematics. Moreover, the newly constituted MEG for 2012 included three U.S. members. This high representation of one country was unprecedented and certainly left the impression that the OECD felt the need for stronger U.S. involvement.

It is worth noting that this U.S. interest in PISA is a relatively new phenomenon. In 2003 I all but begged the National Science Foundation (NSF) to look at disaggregated PISA data to investigate whether students who had gone through the comprehensive reform curricula funded by NSF had significantly different results from other students. These curricula had been aligned directly to the NCTM Standards and thus were geared to improving mathematical literacy. NSF showed no interest at the time. Mostly this was because PISA was not on the U.S. radar in the way that TIMSS was.

However, when the 2003 PISA survey results were announced, the situation changed. Critics of the reform movement and the NCTM Standards were quick to use the mediocre U.S. results as 'proof' that those standards and the curricula that were designed to embody them were a failure. And therein lay an unintended consequence. Up to that point, as I indicated, PISA was far from a U.S. household name. In fact, it had pretty much been dismissed by the right because it measured mathematical literacy, which was in their eyes not as important as mathematical skills. Much more credibility was given to comparisons in curriculum-based assessments, i.e. assessments that are designed around systematic testing of specific mathematical topics taught in schools. But in emphasising the poor results on the PISA survey, PISA itself became emphasised and its importance in the U.S.A. grew from there.

Between 2003 and 2012 we have seen the rise of a new reform movement in the U.S.A. culminating in the CCSSM. And therefore, to some extent the shoe is now on the other foot. When the PISA 2003 results were announced it was clearly unfair to blame the poor U.S. results on the reform curricula at that time, mainly because they had not achieved significant market penetration above the elementary school level. At this time it would be foolish to blame any poor results in the 2012 survey on the policies of the current U.S. administration. But such logic seldom rules in political debates. I think it is safe to predict that any poor results in PISA 2012 will be blamed not on policies of the prior administration but unfairly on the current U.S. government, and possibly on CCSSM despite its very recent implementation.

Given the new-found importance of PISA results in the U.S.A., I believe that there was a move to make PISA a more curriculum-based assessment. The minutes of the first meeting of the 2012 MEG highlight directions from the PISA Secretariat to make the mathematics involved in solving PISA tasks explicit, that authentic tasks were desirable but that these contexts should not constrain the level of mathematical competencies assessed, and that task difficulty should be driven by the mathematics involved and not the complexity of the task context. I believe that the inclusion of three MEG members from the U.S.A. and the involvement of *Achieve* were meant to be steps to move PISA mathematics in accordance with those directions. That a final product evidently acceptable to all stakeholders was achieved is a testament to the MEG members, old and new, to the intellectual leadership of ACER, and to *Achieve* as well.

I found the first meeting of the MEG somewhat tense. But with each subsequent meeting, the MEG came closer and closer to consensus. At our final meeting in Heidelberg in October 2012, MEG member after MEG member spoke to the integrity of the process and the intellectual achievement of the 2012 Mathematics Framework. Given the diversity of the membership and the politically charged atmosphere in which we began, this was no mean feat. I think that it is fair to say that all members believe in and appreciate the importance of promoting mathematical literacy, in the sense of the new Framework definition, throughout the world. We understand that PISA is not a horse race, no matter how the results may be viewed or used. With the change of international contractor for PISA 2015 leading to the exit of ACER from the field and the increased involvement of organisations whose core businesses often involve the commercial publication of textbooks, it is our sincere hope that the essential spirit of PISA can be maintained as it was with the 2012 MEG.

References

Ahmadi, H., & Rafiepour, A. (2013). Mathematics and mobile phone: How students encounter with real world problems. *Roshd Mathematics Educational Journal, 112*, 16–21.

Aunión, J. A. (2007). La educación española retrocede. *El País, 5.12.2007.*

Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effect of international benchmarking in school system performance* (OECD education working papers, N° 71). doi:10.1787/5k9fdfqffr28-en.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. http://www.corestandards.org/965 Math. Accessed 23 Aug 2013.

Instituto de Evaluación. (2011). *Evaluación general de diagnóstico 2010. Educación secundaria obligatoria. Segundo curso Informe de resultados*. Madrid: Ministerio de Educación.

Díaz, P., & Suárez, E. (2010). El nuevo estudio PISA consagra el estancamiento de la educación española. *El Mundo, 7.12.2010*.

Egelund, N. (2008). The value of international comparative studies of achievement: A Danish perspective. *Assessment in Education: Principles, Policy & Practice, 15*(3), 245–251.

Erlangga, D. P. (2012, April). 70 Persen Anak Indonesia Sulit Hidup di Abad 21 (70% Indonesian students realised will find difficult to live in 21st century). *Kompas*. http://edukasi.kompas.com/read/2012/04/04/15495046/70.Persen.Anak.Indonesia.Sulit.Hidup.di.Abad.21. Accessed 23 Aug 2013.

European Communities. (2007). Key competences for lifelong learning: European reference framework. http://ec.europa.eu/dgs/education_culture/publ/pdf/ll-learning/keycomp_en.pdf. Accessed 3 Dec 2013.

Faramarzpour, N., & Rafiepour, A. (2013). What the unwritten rules say about captain's age? *Roshd Mathematics Educational Journal, 112*, 4–9.

Goh, C.-T. (1997). *Shaping our future: Thinking schools, learning nation*. Speech by Prime Minister Goh Chok Tong at the opening of the 7th international conference on thinking on Monday, 2 June 1997, Singapore. http://www.moe.gov.sg/media/speeches/1997/020697.htm. Accessed 23 Aug 2013.

Gooya, Z. (2013). Editor's note. *Roshd Mathematics Educational Journal, 110*, 2–3.

Harwell, M., Post, T., Cutler, A., Maeda, Y., Anderson, E., Norman, K.-W., et al. (2009). The preparation of students from National Science Foundation–funded and commercially developed high school mathematics curricula for their first university mathematics course. *American Educational Research Journal, 46*(1), 203–223. doi:10.3102/0002831208323368.

Hasanpour, M., & Gooya, Z. (2013). Teachers' perspectives about new mathematics textbook in grade nine. *Roshd Mathematics Educational Journal, 111*, 4–9.

Instituto de Evaluación. (2010). *Sistema estatal de indicadores de la educación*. Madrid: Ministerio de Educación.

Kamaliyah, Zulkardi, & Darmawijoyo. (2013). Developing the sixth level of PISA-like mathematics problems for secondary school students. *Indonesian Mathematical Society Journal on Mathematics Education (IndoMS-JME), 4*(1), 9–28.

Karimianzadeh, A., & Rafiepour, A. (2012). Common sense and solving real mathematical problem. *Roshd Mathematics Educational Journal, 107*, 37–44.

Korea Institute of Curriculum and Evaluation (KICE). (2009). *Redesigning elementary and secondary school curriculum for developing future Koreans' core competences*. RRC 2009-10-1. Seoul: KICE.

Kwon, O. N. (2013). Story telling mathematics textbooks: Its development, impact, and prospects. In M. Inprasitha (Ed.), *Proceedings of the 6th East Asia Regional Conference on Mathematics Education (EARCOME6)* (pp. 155–162). Khon Kaen: Center for Research in Mathematics Education.

Legifrance. (2006). Décret n° 2006-830 du 11 juillet 2006, annexe. http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000818367&dateTexte=&categorieLien=id. Accessed 23 Aug 2013.

Lindenskov, L. & Weng, P. (2010). *15 matematikopgaver i PISA: et materiale for matematiklærere og andre med interesse i grundskolens matematikundervisning*. (English: 15 Mathematics tasks in PISA: Material for mathematics teachers and other with interest in mathematics in primary and lower secondary school) (p. 74). Copenhagen: The Danish School of Education, Aarhus University. ISBN: 978-87-7430-198-1.

MENJVA/DGESCO. (2009a). Document d'accompagnement collège, raisonnement et demonstration. http://cache.media.eduscol.education.fr/file/Mathematiques/50/0/doc_acc_clg_raisonnementetdemonstration_223500.pdf. Accessed 23 Aug 2013.

MENJVA/DGESCO (2009b). Banque de problèmes pour les mathématiques. http://cache.media.eduscol.education.fr/file/socle_commun/74/2/Socle_Banque_Mathematiques_College-Problemes_117742.pdf. Accessed 23 Aug 2013.

MENJVA/DGESCO (2011). *Vade-mecum pour la compétence 3*. http://cache.media.eduscol.education.fr/file/DNB/89/2/socle-C3-vade_mecum_166892.pdf. Accessed 23 Aug 2013.

Ministerio de Educación. (2009). *Objetivos Fundamentales y Contenidos Mínimos Obligatorios de la Educación Básica y Media. Actualización 2009*. Santiago: MINEDUC.

Ministerio de Educación y Ciencia. (2006). Ley Orgánica 2/2006, de 3 de mayo, de Educación. BOE, 106, 17158–17207.

Ministry of Education, Denmark. (2003). *Fælles Mål 2003 — Matematik. Faghæfte 12*. (English: Common goals 2003—Mathematics. School Subject Text No. 12). Copenhagen: Ministry of Education

Ministry of Education, Denmark. (2009). *Fælles Mål 2009 — Matematik. Faghæfte 12*. (English: Common goals 2009—Mathematics. School Subject Text No. 12). Copenhagen: Ministry of Education.

Ministry of Education, Science, and Technology (MEST). (2011). *Mathematics curriculum*. Seoul: Ministry of Education, Science, and Technology.

Ministry of Education, Science, and Technology (MEST). (2013). *Mathematics 2*. Seoul: Korea.

Mortimore, P., David-Evans, M., Laukkanen, R., & Valijarvi, J. (2004). *OECD-rapport om grundskolen Danmark* (OECD—Report on primary and lower secondary school in Denmark). Copenhagen: Ministry of Education. Temahæfteserie nr. 5 [English: Thematic serie no 5].

Niss, M. (Ed.). (2002). *Mathematical competencies and the learning of mathematics: The Danish KOM Project*. Roskilde: Roskilde University.

Niss, M. (2003). Mathematical competencies and the learning of mathematics. The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Eds.), *Third Mediterranean conference on mathematics education* (pp. 115–124). Athens: The Hellenic Mathematical Society and Cyprus Mathematical Society.

Nurfuadah, R. N. (2013) Penyebab Indeks Matematika Siswa RI Terendah di Dunia (Reason why Indonesian students have low achievement in the world). *Kampus Okezone*. http://kampus.okezone.com/read/2013/01/08/373/743021/penyebab-indeks-matematika-siswa-ri-terendah-di-dunia. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD). (2005). The definition and selection of key competencies. DeSeCo Executive Summary. http://www.oecd.org/pisa/35070367.pdf. Accessed 11 Oct 2013.

Organisation for Economic Co-operation and Development (OECD). (2006) *PISA released items—mathematics*. OECD Publishing. http://www.oecd.org/pisa/38709418.pdf. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD). (2010a). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. PISA, OECD Publishing. doi:10.1787/9789264062658-en. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD). (2010b). PISA 2009 results: Learning trends. Changes in student performance since 2000 (Volume V). http://dx.doi.org/10.1787/9789264091580-en. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en. Accessed 23 Aug 2013.

Organisation for Economic Co-operation and Development (OECD). (2013b). PISA 2012 released mathematics items. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf. Accessed 8 Oct 2013.

Organisation for Economic Co-operation and Development (OECD). (2003). *The PISA 2003 assessment framework. Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

Rafiepour, A., & Abdolahpour, K. (2013). *Investigating students' modeling competency through grade, gender, and location*. Paper presented at CERME 2013 (8th congress of European Research in Mathematics Education. 6–10 February 2013. Antalya, Turkey.) http://cerme8.metu.edu.tr/wgpapers/wg6_papers.html. Accessed 23 Aug 2013.

Rafiepour, A., & Stacey, K. (2009). Applying a mathematical literacy framework to the Iranian Grade 9 mathematics textbook. In M. Tzekaki, M. Kaldrimidou, & C. Sakonidis (Eds.), *Proceedings of the 33rd conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 433–440). Thessaloniki: PME.

Rafiepour, A., Stacey, K., & Gooya, Z. (2012). Investigating grade nine textbook problems for characteristics related to mathematical literacy. *Mathematics Education Research Journal, 2* (4), 403–421.

Recio, T. (2010). El informe PISA como arma contra el sistema educativo. *El País, 4.12.2010*.

Rico, L. (2011). El estudio PISA y la evaluación de la competencia matemática. *Matematicalia, 7* (1), 1–12.

Rychen, D. S., & Salganik, L. H. (2003). *Key competencies for a successful life and a well-functioning society*. Göttingen: Hogrefe Publishing GmbH. ISBN 9781616762728.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy, 18*(1), 253–286. doi:10.1177/0895904803260042.

Sembiring, R. K., Hoogland, K., & Dolk, M. (2010). *A decade of PMRI in Indonesia*. Utrecht: APS International.

Solar, H., Rojas, F., & Ortiz, A. (2011). *Competencias matemáticas: Una línea de investigación*. Paper presented at the XIII Conferência interamericana de educação matemática, Recife, Brasil. http://www.cimm.ucr.ac.cr/ocs/index.php/xiii_ciaem/xiii_ciaem/paper/view/1867/565. Accessed 23 Aug 2013.

Stacey, K. (2011). The PISA view of mathematical literacy in Indonesia. *Journal of Indonesian Mathematics Society B (Journal on Mathematics Education), 2*(2), 95–126. http://jims-b.org. Accessed 1 Oct 2013.

Unidad Europea de Eurydice. (2002). *Las competencias clave. Un concepto en expansión dentro de la educación general obligatoria*. Madrid: Ministerio de Educación, Ciencia y Cultura.

Wardhani, S., & Rumiati. (2011). Instrumen penilaian hasil belajar matematika SMP: belajar dari PISA dan TIMSS (Instrument evaluation for mathematics achievement in Junior high school: learning from PISA and TIMSS). *Program BERMUTU* (Better education through reformed management and universal teacher upgrading). Yogyakarta: PPPTK Matematika. http://p4tkmatematika.org/file/Bermutu%202011/SMP/4.INSTRUMEN%20PENILAIAN%20HASIL%20BELAJAR%20MATEMATIKA%20...pdf. Accessed 23 Aug 2013.

Widjaja, W. (2011). Toward mathematics literacy in the 21st century: Perspectives from Indonesia. *Southeast Asian Mathematics Education Journal, 1*(1), 75–84.

Zulkardi, Z. (2010). PISA, KTSP and UN. In *Proceedings of the fifteenth national conference on mathematics (Prosiding KNM XV)*. IndoMS and Mathematics Department UNIMA Manado (pp. 53–54). http://eprints.unsri.ac.id/541/. Accessed 23 Aug 2013.

About the Authors

Editors

Kaye Stacey is a Professor Emeritus at the University of Melbourne, Australia, having held the Foundation Chair of Mathematics Education there for 20 years. She is an author of very many books and articles for researchers and also for mathematics teachers at all levels. She worked for four decades as a researcher and teacher educator, training teachers for primary and secondary schools and supervising post-graduate research. Her research interests centre on mathematical thinking and problem solving, student understanding of mathematical concepts, and directions for the mathematics curriculum, especially given the challenges and opportunities that arise from new technologies. She has worked with governments and other agencies on the design of curriculum and assessment to help students learn mathematics in a meaningful way, so that they can think mathematically and use the mathematics they know to solve problems arising from real life. Professor Stacey has been a member of the Australian Research Council College of Experts. She has a Career Research Medal from the Mathematics Education Group of Australasia and a Centenary Medal from the Australian government for outstanding services to mathematics education. She is the mathematics education expert member of the Australian government's advisory group for international assessments and was the Chair of the international Mathematics Expert Group for PISA 2012. Her undergraduate studies were at the University of New South Wales, whilst her doctorate from the University of Oxford is in number theory.

Ross Turner is a Principal Research Fellow at the Australian Council for Educational Research (ACER). He has experience in management, curriculum and assessment, analysis of educational data, and teaching. Mr. Turner contributed significantly to the development of mathematics curriculum and assessment arrangements in the Victorian Certificate of Education (VCE) when it was redeveloped in the late 1980s and early 1990s, and to monitoring and evaluating VCE implementation over subsequent years. Mr. Turner was previously involved in

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7

mathematics education as a secondary teacher for 12 years, and in teacher education programs at tertiary level. Mr. Turner's major role at ACER, since early 2000, has been in management and coordination activities in the Programme for International Student Assessment (PISA), an international research project funded by the OECD. He managed a substantial test development process across three different knowledge domains with test development teams in several countries; led the test development in mathematics; and contributed to other technical aspects of PISA including the methodology used for reporting student achievement. As well as management skills, his technical expertise in the areas of mathematics curriculum and assessment, statistical analysis of performance data, and educational measurement in general are called on regularly as part of his ongoing work at ACER.

Authors

Ferdinando Arzarello is full professor of mathematics at the Department of Mathematics in Turin University. His research deals with mathematics education, and specifically he studies the learning processes of students and the classroom interactions through a semiotic lens. He is author of many scientific papers about these topics. He is also involved in curriculum development programs of the Italian Ministry of Education and in the assessment of students' achievements at different grades. He has been President of the European Society for Research in Mathematics Education (ERME) and currently he is President of the International Commission on Mathematical Instruction (ICMI).

Caroline Bardini was a member of the Mathematics Expert Group for PISA 2012. Her background is in pure mathematics (University of São Paulo, Brazil) and she has specialised in Mathematics Education (Paris 7, France). After a post-doctoral experience in Canada and a European Fellowship in Australia, Caroline worked for 6 years at the Mathematics Department of the Université Montpellier 2, France. Since July 2011, she has been Senior Lecturer in Mathematics Education at the University of Melbourne. Thanks to the international dimension of her career, Caroline is familiar with a wide range of educational systems, which led her to be leader and member of numerous European and international projects as well as chair and member of scientific committees of international conferences. Caroline Bardini's research interests revolve around students' mathematical thinking. She has a keen interest in both examining the impact of technology in mathematics education as well as building bridges between epistemology and mathematics education. Her latest projects revolve around the transition between secondary and tertiary mathematics.

Werner Blum got his Diploma in mathematics in 1969 and his Ph.D. in pure mathematics in 1970, both from the University of Karlsruhe. From 1969 to 1972 he was a lecturer in the Mathematics Department of the University of Karlsruhe, and an assistant professor of mathematics at the University of Kassel until 1975. Since then he has been a full professor of mathematics education (secondary school level) at this university. From 1995 to 2001 he served as the President of the GDM, the maths education society of the German speaking countries. In 2006, he received the Archimedes Award of the MNU, the German association of maths and science teachers.

His research areas include empirical investigations into the teaching and learning of mathematics, for instance on self-regulated mathematics learning and on classroom assessment, and empirical investigations into mathematics teachers' and students' competencies. A main focus of his work is on quality development in mathematics teaching. Among other things, he has been engaged from the beginning in the development of national standards and tests in mathematics for the secondary school level in Germany. He has worked particularly in the area of modelling and applications in mathematics education, for instance as a continuing editor of the series of ICTMA Proceedings and as the editor-in-chief of ICMI Study 14 on modelling and applications in mathematics education, published in 2007.

Leland S. Cogan received undergraduate degrees in psychology and microbiology and his Ph.D. from Michigan State University in Educational Psychology. He is a Senior Researcher in the Center for the Study of Curriculum at Michigan State University and was the Assistant Director for the U.S. Research Center for the Teacher Education Study in Mathematics (U.S. TEDS-M). He has taught courses in educational psychology and educational research methods. Previously he coordinated the data collection and analyses for the Survey of Mathematics and Science Opportunities (SMSO), a multinational project that researched and developed the instruments used in the Third International Mathematics and Science Study (TIMSS). Dr Cogan has co-authored technical reports, articles, and books including *Characterizing Pedagogical Flow*, *Facing the Consequences*, *Why Schools Matter* and *The Preparation Gap: Teacher Education for Middle School Mathematics in Six Countries*. His research interests include evaluation of mathematics and science curricula and the preparation of mathematics and science teachers.

Rossella Garuti has a Ph.D. in mathematics education that was supervised by Maria Giuseppina Bartolini Bussi. She was a mathematics teacher in lower secondary schools from 1984 to 1999. From 1999 to 2007 she was researcher at the IRRE Emilia Romagna (Regional Institute for Educational Research) involved in training of in-service teachers of mathematics. From 2007, she has been principal of a school (kindergarten, elementary school and middle school). Since 2008 she has also been involved in the preparation of the national tests of mathematics from INVALSI (National Institute for the Evaluation of the Education and Training) and mathematics coordinator of the group for Grade VIII. She is author of some research papers in mathematics education and she gave a Regular Lecture at ICME 10.

Toshikazu Ikeda is a professor in mathematics education at the Faculty of Education and Human Sciences of Yokohama National University. He has continuously studied the teaching of mathematical modelling and applications making use of a

Grant-in-aid for Science Research in Japan since 1999. His recent studies are concerned with empirical research on the teaching and learning of mathematical modelling and applications. His focus is mainly on how to support students' discussion to foster types of thinking that will promote mathematical modelling. He obtained the Japan Society of Science Education award in 1999. He was a member of the International Programme Committee for ICMI 14, the International Commission on Mathematical Instruction's study on Applications and Modelling in Mathematics Education and a member of the organising teams for the related topic study groups in ICME 10 (2004) and ICME 11 (2008). Ikeda has also been a member of the International Executive Committee of ICTMA (International Community of Teachers of Mathematical Modelling and Applications) since 2005 and a member of the Mathematics Expert Group for PISA 2012.

Eckhard Klieme is a Full Professor for Educational Sciences at the Johann Wolfgang Goethe-University in Frankfurt/Main and Director of the Department for Educational Quality and Evaluation at the German Institute for International Educational Research (DIPF). From 2004 to 2008 he served as the Managing Director of DIPF. He has a strong background in educational measurement, educational effectiveness, quantitative methods, and comparative studies. Eckhard Klieme graduated from the University of Bonn with master degrees both in mathematics and psychology, and a Ph.D. in psychology. Before joining DIPF, he was a senior researcher at the Institute for Test Development and Talent Research in Bonn (1982–1997), and the Max Planck Institute for Human Development in Berlin (1998–2001). Eckhard Klieme's research focuses on educational effectiveness, school development, and assessment of student competencies. He has been involved in several large-scale assessment programs, both at a national and an international level, including ALL, TIMSS Advanced, TIMSS-Video, PISA, and TALIS. Until 1997, he had major responsibility within the national assessment for medical studies, developing tests of advanced science and mathematics. At DIPF, he directed the German National Assessment of Language Skills (2001–2006), and the indicator-based National Report on Education (2006–2008). He has been involved in the OECD PISA studies since 1998, and is currently Study Director for Questionnaire Development, and chair of the international Questionnaire Expert Group in PISA 2015. Also, he directed research on instructional quality and school effectiveness, including classroom studies on physics education, simulation-based learning, secondary mathematics and early science education, as well as large scale evaluation programs for school improvement.

Fou-Lai Lin is Chair Professor at National Taiwan Normal University. He initiated Taiwanese participation in PISA when he served as the Director of the Department of Science Education, National Science Council. He was appointed the official delegate of Taiwan for the PISA Governing Board Meetings. He was the co-principal investigator of the PISA 2006 study in Taiwan. He makes every effort to promote mathematical literacy for Taiwan. He is also the person in charge for mathematical literacy and reading literacy as the two tests for high achievement assessment. He is the founding Editor-in-Chief for both the *International Journal of*

Science and Mathematics Education and the *Chinese Journal of Science Education* (in Chinese). Meanwhile, he has served as the President for both the International Group for the Psychology of Mathematics Education and the Taiwan Association for Mathematics Education. In addition, he has participated in mathematics test paper development for the College Entrance Examination in Taiwan for more than two decades. His research interests include students' mathematics conceptual understanding, proof and proving, teacher professional development, etc.

Zbigniew Marciniak was born in 1952 in Warsaw, Poland. In 1976 he graduated from the Faculty of Mathematics, Informatics and Mechanics of the University of Warsaw. In 1982 he received a Ph.D. in Mathematics at the Virginia Polytechnic Institute and State University; and in 1997 a post-doctoral degree at the Faculty of Mathematics, Informatics and Mechanics of the University of Warsaw. Professor Marciniak has worked at the University of Warsaw since 1976. From 1996 to 1999 he was the vice dean and from 2000 to 2005 was the dean of the Institute. In the years 2005–2007 he held the post of the President of the State Accreditation Committee. From 2007 to 2009 he was the chairman of the Commission of Didactics in the Committee on Mathematics at the Polish Academy of Sciences. From 2007 to 2009 he was Undersecretary of State at the Ministry of National Education, where he was responsible for defining the main principles in the reform of the education curriculum and the quality of teaching. From 2010 to 2011 he was Undersecretary of State at the Ministry of Science and Higher Education. At present, he is the chairman of the Bologna Team at the Conference of Rectors of the Polish Academic Higher Education Institutions and also a member of the Steering Committee of OECD CERI. Professor Zbigniew Marciniak is an author of more than 30 scientific publications in algebra. He is a member of editorial committees of the periodicals *Delta* and *Algebra and Discrete Mathematics*. For his contribution to supporting mathematically-talented students he was honoured with the Silver Cross of Merit and the Knight's Cross of the Order of Polonia Restituta.

Mogens Niss was trained as a pure mathematician in topological measure theory at the University of Copenhagen and is a full professor of mathematics and mathematics education at Roskilde University, Denmark, which he joined as a member of the founding staff in 1972, after having had a position at the University of Copenhagen 1968–1971. He has been a member of all the PISA Mathematics Expert Groups 1998–2012. In the years 1987–1998, he was a member of the Executive Committee of the International Commission on Mathematical Instruction (ICMI), and the Secretary General of the Commission 1991–1998. He is currently a member of the Education Committee of the European Mathematical Society as well as of the scientific board of the Swedish National Graduate School in Science and Mathematics Education. He is (or has been) a member of several editorial boards of journals, including Educational Studies in Mathematics. Mogens Niss's research deals with mathematics education where his interests and publications focus on mathematical competencies and the justification problem of mathematics education, mathematical modelling and applications, assessment, and the nature of mathematics education research as an academic field. Currently he is involved in designing and implementing an in-service program for upper secondary school mathematics teachers, educating them to help students overcome fundamental learning difficulties in mathematics on the basis of research. In 2012 he was awarded an honorary doctorate at the University of Umeå, Sweden.

Manfred Prenzel Ph.D. is Dean of the TUM School of Education and holds the Susanne Klatten Endowed Chair of Educational Research at the Technische Universitaet Muenchen (TUM). From 1993 he was Professor of Educational Psychology at the University of Regensburg, before he changed in 1997 to the Leibniz-Institute for Science Education (IPN) in Kiel. From 2000 to 2009 he was the Managing Director of IPN. The main topics of his research relate to issues of learning and teaching in different domains (science, mathematics, medicine, economics), especially on motivation and interest, conceptual change, and patterns of teaching and learning. He was the National Project Manager for PISA 2003, PISA 2006 and PISA 2012 in Germany, and a member of the International PISA Science Expert Group from the beginning. Manfred Prenzel is the Director of the Centre for International Student Assessment (ZIB) founded in 2010 by the Federal Ministry of Education and Research (BMBF) and the Standing Conference of the Ministers of Education and Cultural Affairs of the Laender in the Federal Republic of Germany (KMK). This centre unifies the competences of the most important German research institutes in large scale assessment (DIPF, Frankfurt, IPN, Kiel; TUM School of Education, Munich, in strong co-operation with IQB, Berlin). In 2011 Manfred Prenzel was appointed as a member of the German Council of Science and Humanities (Wissenschaftsrat).

Roberto Ricci is the head of the Italian national service of assessment INVALSI. He has a Ph.D. in statistical methodology for scientific research. He mainly deals with the construction of standardised tests for the detection of learning in mathematics. He is the author of several publications in the context of education measurement and investigation techniques of mathematical skills in international surveys (OECD-PISA and IEA).

William H. Schmidt received his undergraduate degree in mathematics from Concordia College in River Forrest, IL and his Ph.D. from the University of Chicago in psychometrics and applied statistics. He carries the title of University Distinguished Professor at Michigan State University and is currently co-director of the Education Policy Center, co-director of the US-China Center for Research and director of the NSF PROM/SE project and holds faculty appointments in the Departments of Statistics and Educational Psychology. Previously he served as National Research Coordinator and Executive Director of the US National Center that oversaw participation of the United States in the IEA sponsored Third International Mathematics and Science Study (TIMSS). Dr. Schmidt has published in numerous journals including the Journal of the American Statistical Association, Journal of Educational Statistics, and the Journal of Educational Measurement. He has co-authored seven books including *Why Schools Matter*. His current writing and research concerns issues of academic content in K-12 schooling, assessment theory and the effects of curriculum on academic achievement. He is also concerned with educational policy related to mathematics, science and testing in general. Dr. Schmidt was awarded the Honorary Doctorate Degree at Concordia University in 1997 and received the1998 Willard Jacobson Lectureship from The New York Academy of Sciences and is a member of the National Academy of Education. In 2009 he was elected in the first group of Fellows in the American Educational Research Association.

Jim Spithill joined ACER in 2010. He has 30 years' experience teaching secondary level mathematics in the government and independent school sectors. At ACER, he has been closely engaged in all aspects of item development for the mathematics component of PISA 2012. This has involved editing items submitted by consortium members; conducting cognitive laboratories with students and implementing the insights obtained; liaising in all aspects of translating items from English to French; reviewing field trial results; assisting in selection and design of final cluster forms; verifying the accuracy of autoscoring procedures for computer-based assessments and contributing to meetings of the PISA 2012 Mathematics Expert Group. Jim has written and reviewed numeracy items for a number of assessment tools for adult and vocational education, and has presented workshops and webinars to technical and further education institutes and registered training organisation personnel about good numeracy assessment items. As well, he has worked on a wide range of ACER projects as a test developer for assessments from primary to senior secondary school levels.

Agnieszka Sułowska has been the PISA Mathematics Head Coder in Poland since 2003. She was also the author of the mathematics part of the Polish national PISA reports 2003–2012. In 2008–2009 she was the Leading Expert at the Ministry of Education for the New Curriculum Project. Since 2007 she has been an external expert at the Central Examination Commission; in particular supervising the process of implementing in Poland the obligatory final high school examination in mathematics in 2007–2010. Currently she is employed as a researcher at the Educational Research Institute.

Dave Tout has over 40 years' experience in the education sector, with most of those years being in vocational, adult and workplace education. He has worked within a range of programs in schools, technical and further education institutes, community education providers, universities, multicultural education services and industry. He also has worked at a state, national and international level in research, curriculum, assessment and materials development. Dave joined ACER in 2008 and has worked on projects including an online Adult Literacy and Numeracy Assessment Tool for the Tertiary Education Commission in New Zealand and the development of online literacy and numeracy assessment tools for both disengaged young people and for adults. Dave took a leading role as test developer and in managing and implementing the item development for PISA 2012. Dave was a member of the Numeracy Expert group for the international Adult Literacy and Lifeskills (ALL) survey and also for the follow up survey, the 2011–2012 Programme in Assessment of Adult Competencies (PIAAC) survey.

Kai-Lin Yang is an Associate Professor in the Department of Mathematics, National Taiwan Normal University. She has served as a mathematics teacher educator for 9 years. Her research interests include reading comprehension of geometry proof and reading strategies for comprehending geometry proof, the assessment of statistical concepts and mathematical modelling, teachers' conceptions of the differences between statistics and mathematics, as well as mathematics textbook analysis based on abstraction. She has published numerous papers in internationally prestigious journals. Dr. Yang's current research deals with students' reading comprehension of geometric construction, the interaction between reading comprehension and problem solving, and teachers' professional development. She also devotes herself to planning and implementing courses for pre-service mathematics teachers and for in-service mathematics teachers.

Index

K. Stacey, R. Turner (eds.), *Assessing Mathematical Literacy*,
DOI 10.1007/978-3-319-10121-7