12th International Congress on Mathematical Education
Program Name XX-YY-zz (pp. abcde-fghij)
8 July – 15 July, 2012, COEX, Seoul, Korea (This part is for LOC use only. Please do not change this part.)

THE INTERNATIONAL ASSESSMENT OF MATHEMATICAL LITERACY: PISA 2012 FRAMEWORK AND ITEMS

Kaye Stacey

University of Melbourne

k.stacey@unimelb.edu.au

The OECD PISA international survey for 2012 is based on a new Framework. This features an improved definition of mathematical literacy; the separate reporting of mathematical processes involved in using mathematics to solve real world problems; and a computer-based component to assess mathematical literacy as it is likely to be encountered in modern workplaces. Issues that arise in the preparation of an assessment for use in many countries around the world will be illustrated with some items and results from the recent international field trial.
KEY WORDS: Mathematical literacy; assessment; comparative studies; computer-based assessment; achievement.

INTRODUCTION

This paper reports on the work done in preparation for the OECD 2012 Programme for International Student Assessment (PISA) survey of mathematical literacy. PISA surveys are conducted every three years. The first was in 2000, so that the 2012 survey is the fifth in the series. As well as making inter-country comparisons and linking achievement data to information on schools and teaching, it is now possible to examine trends in achievement over about a decade. In each cycle, the major focus of the survey rotates through reading literacy, scientific literacy and mathematical literacy. The 2012 survey focuses on mathematical literacy, for the first time since 2003. Because of this focus, a large number of new mathematical literacy items have been created and trialled for 2012, the Framework which specifies the nature of the assessment has been revised and an optional additional computer-based assessment has been developed. The questionnaires for students and schools will emphasise mathematics. In 2012, at least 67 countries (including all 33 of the OECD member countries) will participate in the main survey, with many undertaking the optional components, including computer-based assessment of mathematical literacy (CBAM), general problem solving, financial literacy, parent and teacher questionnaires and a student questionnaire on familiarity with ICT. The first results from the 2012 survey should appear near the end of the year.

There is a great deal of information freely available about PISA, past and present. The official OECD website (http://www.pisa.oecd.org) includes general descriptions of the project, official reports, links to all released items from previous cycles, and secondary analyses of data on topics of interest. The MyPisa website https://mypisa.acer.edu.au hosted by the

Australian Council of Educational Research which leads the International Consortium of contractors, contains or links to copies of student and school questionnaires, national and international reports, research publications, technical manuals and all released items (e.g. http://pisa-sq.acer.edu.au and http://cbasq.acer.edu.au). It is possible to download data bases 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 & Adams, 2007), secondary analyses of PISA data and many reports with a policy or local focus (see, for example, Stacey & Stephens, 2008; Stacey, 2010; Stacey, 2011).

In this paper, I will outline briefly some of the changes and developments for the PISA survey in 2012 that have been the concern of the Mathematics Expert Group (MEG). The outcome of this work is summarised in the new PISA 2012 Mathematics Framework (OECD, 2010), accepted by the PISA Governing Board in 2010, but not yet formally published as it is subject to revision in the light of the results of the main study. The purpose of the Framework is to describe the rationale of the assessment and to define its constructs, describe the components for reporting and specify the nature of the items and the proportion of each type. With the guidance of the Mathematics Expert Group, the Framework was prepared under contract by ACER and Achieve (www.achieve.com) who organised feedback on drafts from over 170 experts in 40 countries.

NEW FRAMEWORK AND NEW EMPHASES

The new Framework has clarified the definition of mathematical literacy, including emphasising the fundamental role that mathematics plays. The intention is to clarify the ideas underpinning mathematical literacy so that they can be more transparently operationalised, whilst retaining have strong continuity with past assessments so that the survey outcomes provide clear evidence of trends in educational outcomes.

Mathematical literacy is still seen as the understanding of mathematics central to a young person's preparedness for life in modern society, from simple everyday activities to preparing for a professional role. Even more than when the PISA project was first devised, a growing proportion of problem situations encountered in work and life require some level of understanding of mathematics, mathematical reasoning and use of tools with a mathematical aspect. The notion of mathematical modelling (especially through de Lange's theorisation of mathematisation) was the cornerstone of the PISA Framework for mathematics from the start (OECD, 2003) and it remains so. Now it has been more explicitly drawn into the new definition, with explicit reference to the component processes of modelling namely (i) formulating real world problems mathematically, (ii) employing mathematics to solve the mathematically formulated problem and then (iii) interpreting and evaluating the mathematical results in real world terms. The new 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, 2010, p. 4)

The Framework directly addresses the misconception that mathematical literacy is synonymous with minimal, or low-level, knowledge and skills. It proposes a continuum of mathematical literacy from low levels to high levels, not a cut-off point above which one is mathematically literate. Mathematical literacy is intended as a construct applicable to all ages and levels of expertise. However, assessment of 15-year olds must take into account relevant characteristics of these students, including the mathematical content they are likely to know.

The various components of the Framework are illustrated in Figure 1. Consistent with the definition of mathematical literacy, PISA items are almost all set in a real world context (indicated by the outside box) although there is variation in the degree to which solving the problem requires engagement with the context (whether the context functions as a border, wrapper or tapestry in the terms of Stillman (1998) or as first to third order context in the terms of de Lange (1987)). The personal, societal, occupational and scientific *context categories* point to the areas of life where mathematical literacy is required (from the everyday to the professional) and from which item contexts will be drawn. The Framework specifies the percentage of PISA items to be in these (and other) categories, to ensure a balance in the assessment. The context categories are renamed (a simplification) from the previous Framework, and there is some minor regrouping. The few intra-mathematical items are in the scientific category.

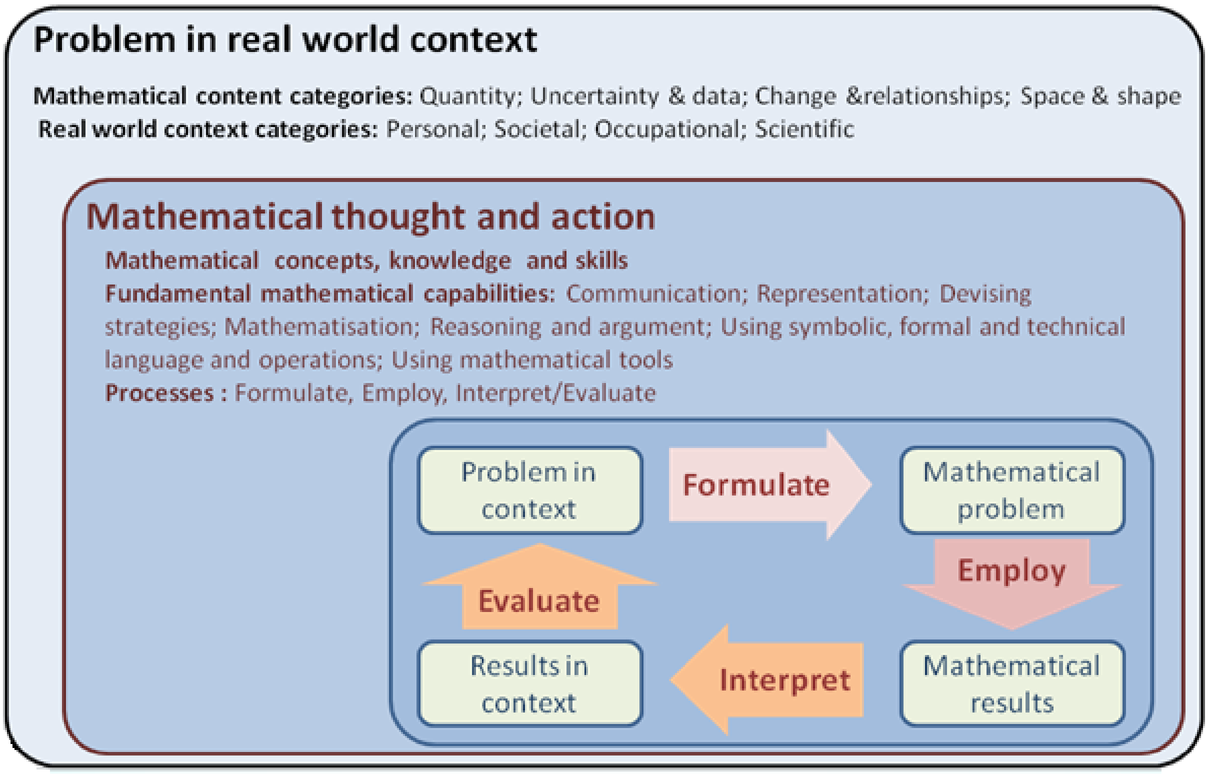


Figure 1. A model of mathematical literacy in practice.

Still in the outside "Problem in real world context" box, PISA items can also be categorised according to the 4 major mathematical themes which underlie the situations and structures involved. In the 2003 Framework, these categories were called the 'overarching ideas', but they are now called *mathematical content categories*, although the construct is the same. There has been minor renaming, so that 'Uncertainty' is now 'Uncertainty and data' in order to more formally recognise that dealing directly with data (not uncertainty about data) is a key ability for citizens making judgements and decisions. Feedback on the draft Framework highlighted the difficulty that some experts around the world had with this point, so hopefully the renaming will communicate more clearly. The content categories classify the items according to major mathematical structure and themes inherent in the mathematical core of the problem. These are the underlying ideas which have inspired the various branches of mathematics. For example, change is a theme encountered in many contexts, which can be handled with rates, functions or calculus. The *Space and shape* description describes geometry and measurement as foundational to activities such as perspective drawing, creating and reading maps, transforming shapes with or without technology, dealing with images of three dimensional scenes and representing objects and shapes.

An important addition to the Framework is a move towards a more explicit description of mathematical content that is appropriate for assessment of 15 year old students. Expert advice was moderated by a survey conducted by Achieve (www.achieve.com) of published curricula from 11 countries. The aim was to clarify what 15 year olds will have had the opportunity to learn and also what countries deem realistic and important preparation as students approaching entry into the workplace or higher education. As the examples below show, the content is described in broad terms, in contrast to the careful curriculum analysis used in the TIMSS tests. The list is intended to be a guide for item writers, rather than prescriptive. It is also recognised that there is not a one-to-one relationship between mathematical topics and the content categories.

> "*Algebraic expressions*: Verbal interpretation of and manipulation with algebraic expressions, involving numbers, symbols, arithmetic operations, powers and simple roots. *Equations and inequalities*: Linear and related equations and inequalities, simple second-degree equations, and analytic and non-analytic solution methods". (OECD, 2010, p 13)

FUNDAMENTAL MATHEMATICAL CAPABILITIES

Dealing with a problem in the PISA assessment necessarily involves mathematical thought and action. As depicted in the middle box of Figure 1, this is conceptualised as having three components. The first is *Mathematical concepts, knowledge and skills*: the knowledge base, consisting of mathematical concepts, known facts, and skills in performing mathematical actions (the topics described above). Second, activating the knowledge base, are the seven *fundamental mathematical capabilities* (FMC) for mathematical action. These capabilities are derived from the mathematical competencies of the 2003 Framework, and indeed the original PISA Framework (OECD, 1999) which were based on work done by Mogens Niss (a longstanding member of the MEG) and colleagues (Niss, 1999; Niss, 2003; Niss and Højgaard, 2011). In the new Framework, these have been renamed to fit better with terminology in other OECD assessments, and simplified based on empirical work conducted with the Mathematics Expert Group (Adams, 2012; Turner, 2009; Turner, Dossey, Blum & Niss, in press; Turner, 2012). Given the importance of new technology for doing mathematics, and the new possibilities of the CBAM, *Using mathematical tools* is an additional FMC.

The FMC describe the various processes that are increasingly seen as central to an individual's understanding of mathematical ideas and capacity to apply his or her mathematical knowledge. Evidence of this recognition can be seen in the formal curriculum statements of various educational jurisdictions around the world. Empirical evidence of the centrality and importance of these capabilities to mathematical performance would be established if a strong relationship can be found between item difficulty (measured empirically) and ratings of items on the capabilities. Consortium staff and members of the MEG have been engaged in this work; describing four graduated levels of operation of each of the capabilities and judging the extent to which successfully answering PISA questions demands their activation. It is recognised that the chosen capabilities overlap to some extent, and that they frequently operate in concert and interact with each other; nevertheless the rating procedure has been to treat each competency as distinctly as possible. Early statistical work reduced the number of (then) competencies to the present number and examined inter-rater consistency, leading to improved level descriptions in an iterative process (Turner, 2012; Turner et al, in press). Ongoing work will test the power of the ratings according to the refined levels to predict item difficulty in the new PISA study (Adams, 2012). It is expected that the completed scheme will have wide application in understanding the capabilities that underpin mathematical literacy and creative mathematical activity more generally, as a useful tool for item developers in many situations, and as a guide for teachers to demonstrate both capabilities to emphasise and levels of development through which they students may progress. The FMC with their level descriptions are used to describe the 6 levels of the proficiency scale, which (along with point scores) are used for reporting PISA performance.

NEW REPORTING OF PROCESSES

The third part of mathematical thought and action are the processes of solving problems involving mathematical literacy. In previous cycles, mathematical literacy has been reported on one overall proficiency scale and also by the 'overarching ideas' (now the mathematical content categories). However in 2012 an attempt will be made to report also against the processes of using mathematics to solve real world problems. Figure 1 lists the processes and the inner box sketches a very simplified description of how they are related to mathematical literacy. A real world problem first needs to be transformed into a mathematical problem: This is the *formulate* process (full name *Formulating situations mathematically*.), indicated by the top arrow. The mathematical problem is solved by the *employ* process (full name *Employing mathematical concepts, facts, procedures, and reasoning)*. This is indicated by the rightmost arrow. The mathematical results that are produced then need to be translated into real world terms and judged for their adequacy. If inadequate, the problem situation may need to be reformulated. This is the *interpret process* (full name *Interpreting, applying and* *evaluating mathematical outcomes)* which is indicated by the bottom and left arrows of the mathematical modelling cycle in the diagram.

The intention is that in 2012, PISA mathematical literacy results will be reported as overall score, scores for each of the four mathematical content categories, and (new) scores for each of the three processes. It is hoped that this additional reporting structure will provide useful and policy-relevant results. The scores on the *formulate* scale should show how effectively students are able to recognise and identify opportunities to use mathematics in problem situations and then provide the necessary mathematical structure needed to formulate that contextualised problem into a mathematical form. The *employ* scale should indicate how well students are able to perform computations and manipulations and apply the concepts and facts they know to arrive at a mathematical solution to a problem formulated mathematically. The *interpret* scale should indicate how effectively students are able to reflect upon mathematical solutions or conclusions, interpret them in the context of a real-world problem, and evaluate whether the results or conclusions are reasonable. Students' facility at applying mathematics to problems and situations is dependent on all three of these processes, and an understanding of their effectiveness in each process can help inform both policy-level discussions and decisions being made closer to the classroom level.

The field trial results have been analysed by Ray Adams (personal communication). It shows that there are very high correlations between the student scores on the three process scales, especially for the paper-based assessment. This indicates a redundancy in the information. However, they are comparable to the correlations between component scales for other domains in previous field trials, and these have proved to be useful for country comparisons and comparisons of identified subgroups of students. In part, the high correlations may be explained by the difficulty of allocating items to only one of the processes (multiple allocation is not advisable for the data processing). Additionally, there is a tension between writing items with strong face validity for mathematical literacy (which tend to involve all stages of the modelling cycle) and writing items which can be unequivocally allocated to one of the three processes. Another factor is that the significant time limits on answering a PISA question means that items often hone in on a specific aspect, but in such a case it can be unclear whether the student understands this to be putting them at the beginning (*formulate*) or the end (*interpret*) of the modelling process. Despite these inevitable 'boundary disputes' for categorisation the first analysis shows some country-level spread across processes.. The first glimpse of the field trial data also shows potentially interesting differences in the performance of countries on the process scales derived from paper-based tests and computer-based items. The field study showed that items classified as *formulate* (i.e. items where this was judged to be the main source of cognitive demand) tended to be harder than other items. This observation concurs with the lament of many teachers: "my students cannot do word problems". For the test as a whole, items should be selected so that there is a good range of difficulty for items of every type, so easy *formulate* items were in demand when selecting items.

ITEM DEVELOPMENT

The 2012 mathematics assessment required 72 new items to be used with 36 link items from earlier surveys to calculate trends. New items, proposed by teams around the world, went into a large pool which was approximately halved for the field trial and halved again for the main study in the light of empirical results. Before selection for the field trial, items were subject to intensive scrutiny by the MEG, by external experts organised by Achieve, and by the national teams in every country. Items were generally very well received by these reviewers, indicating the substantial imagination and expertise of the writing teams. The field trial showed that all but a few of the items had very good statistical properties, so there were good options for making a final selection that meets the Framework's multiple criteria of balance across the 4 context, 4 content and 3 process categories, with items of an appropriate range of difficulty in each cell of the matrix.

Although many items are publically released from previous cycles (see links above), I will illustrate some of the points about item development three new items of one unit, which went into the field trial in 2011 but were not selected for the main study. There are many reasons for not selecting an item. These include differential performance in countries (likely to be due to cultural or linguistic factors); lack of support from national teams; statistical anomalies at the field trial (e.g. poor discrimination, worrying gender or country differences, too hard or too easy); occasionally a logical or mathematical content flaw identified late in the process; or simply because it is surplus to requirements.

For the 2012 survey, as in the past, a disproportionately large number of the harder items were not selected, because an efficient test should bunch the difficulty of the items near the ability of most of the students, and the items in the field trial tended to be relatively hard for the students. It is a waste of time and money to include in PISA surveys, items which only a small percentage of students can do, since they do not contribute meaningfully to the reported results. Item writers, even those who are very experienced, tend to create relatively too many items which have very low success rates.

Many people (myself included) who examine the mathematics in the PISA items feel that it is not of a 'high standard' for Year 10 students, and there was an attempt in the field trial to ensure that this concern was addressed. However, in evaluating the nature and level of mathematics in the PISA items, it is important for everyone concerned to focus on whether the tested mathematics is important – whether knowing it makes a significant contribution to students' mathematical literacy as they require it in their lives. A trap is to confuse significant mathematics with hard mathematics. Percentage calculations and the associated proportional reasoning, for example, are clearly of significance in mathematical literacy, although not seen as hard for 15 year olds. That being said, it is still of concern that the success rates in PISA items are in general surprisingly low, even in countries that are highly ranked. There is a great deal of room for improvement in mathematical achievement even in countries that do well. One hypothesis is that the difficulty of PISA items stems from the mathematical literacy focus, presenting them in contexts, but potentially it may be caused by the conditions of taking the test. This is not known, but use of the polished scheme for rating the influence of the capabilities may provide a research tool to help in understanding this.

In order to match item difficulty better with student ability, some countries which have had low achievement in the past have chosen to test in 2012 with an item pool which includes relatively more easy items. This will enable more reliable measures to be made at the lower end of the achievement scale, and so enhance the usefulness of their results for policy development. Complex statistical procedures, using a Rasch model place individual students on one common scale, despite the fact that they do different booklets (subsets) of items, and then combine student level data to create comparable country scores. Because of the complex processing involved, individual students cannot be given meaningful feedback on their performance immediately after completing the test.

Example: Chinese Lamp (PM999)

This example illustrates a new unit containing three items from the field trial that, because of unpredicted poor statistical performance (low discrimination), will not be used in the main study and hence can now be released. *Chinese Lamp* was a unit that was well regarded by all the reviewers, and so can be used to illustrate the nature of mathematical literacy items, and the processes of item development and categorisation. Its poor statistical performance was not predicted, showing the importance of the field trial. The *Chinese Lamp* items were generally approved by the national PISA teams, with the average "priority for inclusion" at 3.5, 3.4 and 3.2 on a 5 point scale across countries. Item statistics reported below are based on over 6000 students from the 25 OECD countries whose results had been fully analysed at the time when selection of items for the main study was undertaken. The sampling for the field trial is not strictly controlled as it is in the main study, but previous experience indicates the field trial results are a good guide.

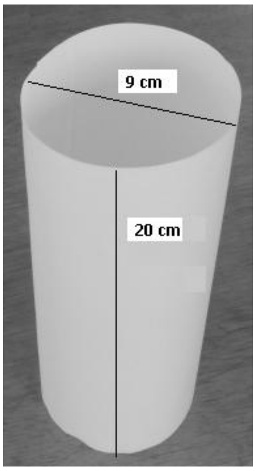
The three items in the unit *Chinese Lamp* (PM999) shown in Figure 2 (stem) and Figure 3 (items) belong to the Shape and Space content category in the personal context category. With only a minor change of the cover story, the unit could be altered to belong to the occupational context category. The first two items (Questions 1 and 2 in Figure 3) were allocated to the mathematical process category 'formulate' and Question 3 was allocated to 'interpret'. This will be discussed below.

Figure 2 retains the notes that assist translators to distinguish those parts of the text which are intended to be common language descriptions (in this case both 'outside angle' and the description of the folding) from those which have specific mathematical terminology which needs to be translated precisely. The PISA project has very high standards of translation, with multiple procedures requiring translating into the third language from very carefully matched originals in both English and French, then comparing the third language product as well as back and cross translations. Although mathematics is often thought of as a universal language, PISA items present many translation challenges. A major source of the challenge is the relatively informal language that is authentically used in mathematical literacy items to describe real world contexts contrasting with formal mathematical language. The *Chinese Lamp* provides examples of this ('triangle shapes', 'sheets are 20 cm high', 'outside angle'). Differences also come in mathematical descriptions. English, for example, has many ways of expressing a ratio: "three times longer than" (Figure 3), the more correct 'three times as long as"), "a ratio of 3:1" or 'triple'. Some languages do not have such a wide range of alternatives and so in some cases less translation from colloquial language to formal mathematics may be required by students undertaking the assessment in some languages than others. This is guarded against in the field trial by checking for unexpected country differences.

Chinese Lamp

Mira is organising a party at her house and would like to make Chinese lamps. Each lamp is made of two sheets of special paper: an internal sheet rolled in a cylinder (Picture 1) and a second sheet folded to create 12 triangles (Picture 2). The diameter of the internal paper cylinder is 9 cm and its height is 20 cm. The triangle shapes in Picture 2, that are created by the folding, form approximately equilateral triangles.

Translation Note: In this unit please retain metric units throughout.
Translation Note: Please adapt the terminology if you have a specific expression for this kind of folding. For example, in FRE it is "feuille pliée en accordéon".



Picture 1

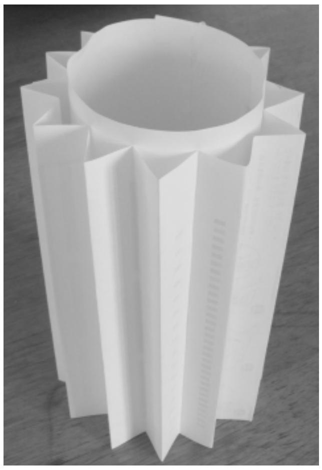

Picture 2

Figure 2: Chinese Lamp (PM999Q01) unit stem. Copyright OECD.

Question 1 (see Figure 3) requires spatial reasoning to identify how the paper will be folded to make the cylinder and to link the length of paper to the perimeter of a circle through the cross-sectional view, and then calculating the perimeter approximately. This item is allocated to the 'formulate' process (rather than 'employ') because it was judged that the main cognitive demand came from the spatial reasoning rather than from the calculation of the approximate circumference from the diameter. This item was a middle difficulty item in the field trial with 45.63% of students correct. Another 35% of students were approximately equally spread between the answers of (a) 20 cm and (c) 40 cm, and about 7% selected each of (d) 50 cm and (e) 60 cm. This item had low discrimination. One indicator of this is that the 7% of students who selected (e) had higher average proficiency (as judged by the rest of the test) than students selecting the correct answer. Their ability was considerably higher than students who selected (d), which is an indicator that there may be quite different reasons for making these two choices. Because the aim of the PISA survey is to measure students' mathematical literacy as effectively and efficiently as possible, items of low discrimination are normally discarded. In a school assessment or public examination, there are other goals of assessment to consider – such as testing across a curriculum and illustrating to teachers important aspects of mathematics in accordance with the MSEB's "Learning Principle" (Stacey & Wiliam, to appear). However, in an international assessment which is kept secure, the considerations are different. There was a moderate gender difference favouring girls on this item and also on Question 2, but not on Question 3 where the sexes performed equally.

Question 1

In Mira's favourite hobby shop they sell several sheet sizes. The sheets are all 20 cm high but have different lengths.

Which of the following is the smallest length Mira could buy in order to make the paper cylinder (see Picture 1)? (Note that at least 0.5 cm extra is needed for gluing.)

a) 20 cm
b) 30 cm
c) 40 cm
d) 50 cm
e) 60 cm

Question 2

In a finished Chinese lamp (see Picture 2), how many times longer is the folded sheet of paper compared to the length of the internal sheet rolled in a cylinder?

a) Approximately 1.5 times
b) Approximately 2 times
c) Approximately 3 times
d) Approximately 12 times

Question 3

Mira wants to create another Chinese lamp in a similar style.

Which of the following changes to the lamp shown in Picture 2 will affect the length of the folded sheet of paper?

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

a) Keep the same size paper cylinder, change the size of the outside angle in each of the 12 triangles of the folded sheet of paper from about 60^{o} to about 30^{o}. Yes / No
b) Keep the same size paper cylinder, increase the number of equilateral triangles of the folded sheet of paper from 12 to 20. Yes / No
c) Change the diameter of the internal paper cylinder, keep the folded sheet of paper with 12 equilateral triangles. Yes / No

Translation Note: Please do not translate "outside angle" as "external angle" as this term has a different and precise mathematical meaning.

Figure 3: Chinese Lamp questions with correct answers underlined. Copyright OECD.

Question 2 (see Figure 3) again requires considerable spatial reasoning, to identify the importance of the view from the top and to see the 12 nearly equilateral triangles, and then to see that the circumference of the cylinder is made up of one side of each 'equilateral triangle'. The folded paper is therefore close to twice the length of the circumference of the cylinder. Again this is categorised as 'formulate', identifying the spatial reasoning as a more important source of cognitive demand than employing knowledge of equal side length of equilateral triangles. The percentages of students correct (option (b)) was 42.22%, whilst about 20% of students chose each of options (a) and (c). Only 9.71% chose option (d), where the 12 times in the response superficially matches the 12 triangles in the diagram. This item also has low discrimination. Perhaps this is because it can be solved by the relatively complex spatial reasoning above, or by simpler methods such as direct estimation of comparative length from the diagram, or even by measuring the folds in the picture with a ruler (about 1 cm each on my copy).

Question 3 (see Figure 3) is a complex multiple choice item. This format significantly reduces the chance of randomly choosing a correct response, in this case to 1 in 8. As with most complex multiple choice items, a response to this item was only scored correct if all parts are right. In total, 24.21% of students were correct (i.e. all parts right), and about 60% were right on one or two parts, but scored no credit for the whole item. Again, the item has low discrimination. For example, the correlation of the score on this item with the score on the rest of the test was 0.06. Statistics for individual parts are not available. I guess that the second part was the most difficult. This item was classified as 'interpret' but arguments could also be made for a classification as 'formulate'. It is important to note that the 'interpret' process is not about interpreting the meaning of the statements (i.e. receptive communication) or reading mathematical representations (e.g. understanding a graph), but about putting mathematical outcomes into real world terms, and evaluating the adequacy of solutions. It refers to the shift back from the mathematical world into the real world.

COMPUTER-BASED ASSESSMENT OF MATHEMATICS (CBAM)

A major initiative for the 2012 survey is the introduction of the optional computer-based assessment of mathematics CBAM. This follows the development of assessment of electronic reading (2009) and computer-based assessment of scientific literacy (2006). 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.

Behind the introduction of all of these computer-based assessments is a long term view to shift all PISA assessment from paper-based to computer-based. Developing expertise and infrastructure is therefore valuable. However, there are specific reasons why computer-based assessment is important for mathematics and for mathematical literacy. First, computers are now so commonly used in the workplace and in everyday life that a level of competency in mathematical literacy in the 21st century includes using computers. Hoyles, Wolf, Molyneux-Hodgson, & Kent (2002) note that mathematical literacy in the workplace is now completely intertwined with computer literacy, at all levels of the workforce. Doing mathematics with the assistance of a computer is now part of mathematical literacy.

A second consideration is that the computer provides a range of opportunities for designers to write test items that are more interactive, authentic and engaging, and which may move mathematics assessment away from the current strong reliance on verbal stimuli and responses, enabling different student abilities to be tapped (Stacey & Wiliam, in press). These opportunities include using new item formats (e.g., drag-and-drop or hotspots), presenting students with flexible access to real-world data (such as a large, sortable dataset), moving stimuli and simulations, representations of three-dimensional objects that can be rotated, or to use colour and graphics to make the assessment more engaging. The effect of the latter apparently trivial factor has been immediately recognised in the cognitive laboratories where individual students are interviewed solving the items during initial item development. By permitting a wider range of response types, CBAM could give a more rounded picture of mathematical literacy.

Future PISA cycles may feature more sophisticated computer-based items, as developers and item writers become more fully immersed in computer-based assessment and the infrastructure develops. Indeed, PISA 2012 represents only a starting point for the possibilities of the computer-based assessment of mathematics. In the future, I expect to see assessment of proficiencies the interface of mathematics and ICT. Examples would include making a chart from data, including from a table of values, (e.g., pie chart, bar chart, line graph) or producing graphs of functions to answer questions about them. In the present assessment, due to expectations of what 15 year olds around the world may be able to manage now, using generic mathematical tools does not go beyond the on-screen calculator.

A key challenge is to distinguish the mathematical and mathematics-with-ICT demands of a PISA computer-based item from demands unrelated to mathematical proficiency such as using a mouse, understanding basic conventions such as clicking on arrows to move to a new screen. The latter category of demand needs to be minimised.

Examples from the CBAM field trial

In the field trial, computer-based assessments were administered to 39970 students in 52 locales (a locale is a language by country combination) across 43 countries in 30 languages, with considerable operational success. For CBAM, the field trial tested 86 items, from which the final pool of 45 items were to be selected. Each student will do about a quarter of these in the main study, with data on time taken being used to select the subsets.

Four sample units (translated into 14 languages) are available on the website (cbasq.acer.edu.au). Please note that these items will not be included in the main study, either because of statistical anomalies from the field trial, a flaw in the item, or because they are surplus to requirements.

The two sample units *Graphs* (CM010) and *Car cost calculator* (CM013) provide examples of items that could equally well be paper-based items, items where computer presentation enables a new response type, and items which require use of some computational power of the computer. The first item of the unit *Graphs* asks students, in multiple choice format, to identify a real world situation that could relate to a given bar graph. In this case, the bar graph shows regular cycling, and the correct choice is maximum monthly temperature of a city, rather than to temperature of a cup of coffee, weight of a baby, or diminishing coal reserves. There were 49.53% correct in the 15 OECD countries of the CBAM field trial. This could be a paper-based item: it does not need computer presentation.

The second and third items take advantage of the computer presentation for enhanced responses and automated marking. Here the computer-based format offers both new opportunities and increased convenience. Students have to construct bar graphs, by dragging prepared bars onto the graph to match a giving situation. Item 3 is shown in Figure 4, partially completed with three bars already dragged onto the axes. The instruction is “Drag and position each of the bars onto the Time axis to show how Jenny's yearly income changed over the 10-year period.” Answering in paper-based format would be very impractical, because students would have to draw bars of precise heights and would be likely to need several trials. Also scoring by hand would be time consuming. Mathematically, this is a difficult item, because the information about the constant annual increase needs to be carefully considered, but using qualitative reasoning about the constant gradient property of constant increase. Only 10.57% of students were correct, taking an average of 167 seconds to respond. The capacity to track additional characteristics of computer-based responses, such as response time or number of attempts, can provide researchers and test developers with additional insights into students’ performance (Stacey & Wiliam, in press). This capacity will grow as infrastructure for computer-based tests is refined in the future. One reason why this item has been released is that it had low discrimination in 6 of 15 OECD countries in the field trial.

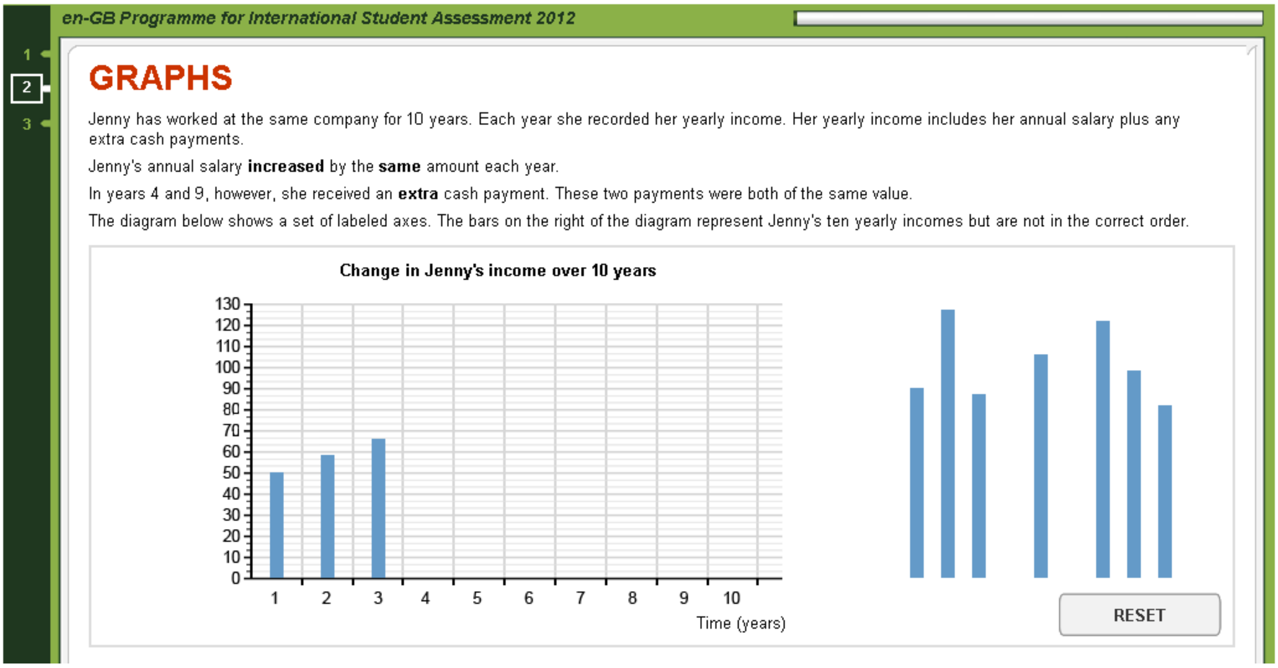
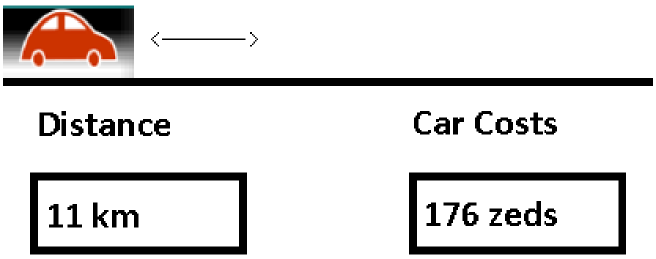


Figure 4. Partially completed Item 3 of *Graphs* (CM010Q03). Copyright OECD.

Some of the computer-based items have students directly use the computational power of a computer, both for calculation and exploring the mathematical structure inherent in a real world context. Items requiring use of specific mathematically-able software (e.g. to program a spreadsheet, or use a generic tool to plot of graph) have not been used at this early stage, because of the very varied abilities of students around the world to use it. However, there are a great number of user-friendly calculators available on websites around the world to assist consumers, so mimicking this provides an opportunity for mathematical thinking in a very realistic setting.

The unit *Car Cost Calculator* (CM013) (Figure 5) provides an example of this. Dragging the car causes the distance and monthly cost of travelling to work and back by car to change. Figure 5 shows a simplified interface. The first item requires direct use of the car cost calculator followed by other calculation (perhaps with an ordinary or on-screen calculator). It asks what percentage of car travel cost would be saved by buying a monthly transport ticket for a person travelling 15km (36.46% correct). The second item could as well be paper-based, asking students to select a formula for working out petrol costs as a function of distance to work, given appropriate data (18.80% correct, with low discrimination in most countries). The third item (see Figure 5) requires obtaining sensible data from the car calculator to work with algebra. On average, the National Program Managers of the participating countries expressed approval for all of these items to be included (average scores between 3.4 and 3.9 on a 5 point scale) which indicates that they judged the mathematical content and real world setting to both be appropriate for their students. The poor discrimination was not predicted by the expert teams.

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.



Question 3. (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: $\boldsymbol{C = 6d + b}$

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

Figure 5. Illustrated item CM013Q03 – drag the car and the distance and car costs change.

STUDENT QUESTIONNAIRE

In 2012, the questionnaires for students and schools (20 to 30 minutes) will contain many general items about school as well as items that focus on mathematics, prepared by the PISA questionnaire expert group and consortium staff. There will be questions about the mathematics learning environment, students' opportunity to learn mathematical literacy at school, their interest in mathematics and their willingness to engage in it. Responses will be related to achievement scores. *Opportunity to Learn* items relate to student experience with applied mathematics problems, familiarity with mathematical concepts by name, and student experience in class or tests with PISA style items. *Interest in mathematics* relates to present and future activity: mathematics at school, perceived usefulness in real life and intentions to undertake further study and/or mathematics-oriented careers. *Willingness to engage in mathematics* taps into emotions of enjoyment, confidence and (lack of) mathematics anxiety, and the self-related beliefs of self-concept and self-efficacy. There is international concern about interest and willingness because of a decline in the percentage of students who are choosing mathematics-related future studies, whereas at the same time there is a growing need for graduates from these areas. A recent analysis of the subsequent progress of young Australians who scored poorly on PISA at age 15 found that those who "recognise the value of mathematics for their future success are more likely to achieve this success, and that includes being happy with many aspects of their personal lives as well as their futures and careers" (Thomson & Hillman, 2010, p. 31). The study recommends that a school focus on the practical applications of mathematics may go some way to improving the outlook for these low-achieving students.

CONCLUSION

The aim of this paper was to outline some aspects of the PISA 2012 mathematical literacy assessment, and to highlight some of the theoretical and practical changes that have been introduced. The new Framework has strong continuity of the past, so that trends can be measured, but it has also addressed earlier critiques, by some clarification and simplification, by emphasising the centrality of mathematics more strongly, and placing the multiple components of the Framework into a holistic picture. A major strength of PISA lies in the rigorous procedures for item development, review, trialling and selection, for translation and coding (not addressed above), for sampling in the main survey and expert data analysis for reporting the results. It is hoped that the additional reporting categories (the three processes) of PISA 2012 will enhance the usefulness of the results for public policy development and provide further insights into how the nature of the mathematics provision in schools affects mathematical literacy. On-going work in rating the influence of the fundamental mathematical capabilities on the total cognitive demand of items (checked against empirical measures as they become available) may prove useful to item developers for PISA and other assessments, and provide new insights into what makes mathematical literacy items difficult. The new initiative of computer-based assessment of mathematics is an important one, which opens up new avenues for probing mathematical literacy with and without technological assistance. I look forward to the release of the first results at the end of 2012.

Acknowledgement

I acknowledge with thanks the contributions of Ross Turner, Ray Adams, other Consortium staff, Achieve staff and the ten members of the international Mathematics Expert Group. The author is Chair of the Mathematics Expert Group for the PISA 2012 survey. Opinions expressed in this paper are those of the author alone.

References

Adams, R. (2012). Some Drivers of Test Item Difficulty in Mathematics: An Analysis of the Competency Rubric. Paper presented at AERA Annual Meeting, April 13- 17, 2012 Vancouver, Canada

De Lange, J. (1987). Mathematics – insight and meaning. Utrecht: Rijksuniversiteit Utrecht.

Hoyles, C., Wolf, A., Molyneux-Hodgson, S. & Kent , P. (2002). *Mathematical Skills in the Workplace. Final Report to the Science, Technology and Mathematics Council*. London: University of London and Science, Technology and Mathematics Council.

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

Niss, M. (2003). Mathematical Competencies and the learning of mathematics: The Danish KOM Project. In Gagatsis, A. & Papastavridis, S. (Eds.). *3rd Mediterranean Conference on Mathematics Education* (pp. 110-124). Athens, Greece: The Hellenic Mathematical Society and Cyprus Mathematical Society.

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.

OECD (2003). *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD Publications.

OECD (2009). *PISA: Take the test.* Paris: OECD Publications. http://www.oecd.org/dataoecd/47/23/41943106.pdf

OECD (2010). *PISA 2012 Mathematics Framework*. Paris: OECD Publications http://www.oecd.org/dataoecd/8/38/46961598.pdf

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. http://www.dest.gov.au/sectors/school_education/publications_resources/schooling_issues_digest/

Stacey, K. & Wiliam, D. (in press). Technology and Assessment in Mathematics. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third International Handbook of Mathematics Education*. Springer.

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- Journal on Mathematics Education*, 2(2).

Stillman, G. (1998). The emperor's new clothes? Teaching and assessment of mathematical applications at the senior secondary level. In P. Galbraith, W. Blum, G. Booker & I. Huntley (Eds.), *Mathematical modelling: Teaching and assessment in a technology rich world.* (pp. 243-253). Chichester: Horwood Publishing.

Thomson, S. & Hillman, K. (2010). *Against the odds: Influences on the post-school success of 'low performers'*. Melbourne: Australian Council for Educational Research http://research.acer.edu.au/transitions_misc/7

Turner, R. & Adams, R. (2007). The Programme for International Assessment: An Overview. *Journal of Applied Measurement*, 8(3), 237-248.

Turner, R. (2009). Using Mathematical Competencies to Predict Item Difficulty in PISA: A MEG Study 2003-2009. Paper presented at PISA Research Conference, Kiel 2009. http://www.pisaresconf09.org/user_uploads/files/symposium/mathematics/Turner.pdf

Turner, R. (2012). Some Drivers of Test Item Difficulty in Mathematics: An Item Analysis Rubric. Paper presented at AERA Annual Meeting, April 13-17, 2012 Vancouver, Canada

Turner, R., Dossey, J., Blum, W., & Niss, M. (in press). Using mathematical competencies to predict item difficulty in PISA. In M. Prenzel, M. Kobarg, K. Schöps, &S. Rönnebeck. (Eds.), *Research in the context of the Programme for International Student Assessment*.