PIAAC Numeracy Task Complexity Schema: Factors that impact on item difficulty

Dave Tout, Iddo Gal, with Mieke van Groenestijn, Myrna Manly and Mary Jane Schmitt

Australian Council *for* Educational Research

ACER

Published by the Australian Council for Educational Research

19 Prospect Hill Road

Camberwell, Victoria, 3124, Australia

Ph: +61 3 9277 5555

ISBN: 978-1-74286-609-3

Authors

Dave Tout, Senior Research Fellow, Numeracy & Mathematics, Tertiary and Vocational Assessment Services, Australian Council for Educational Research (ACER), Australia

Iddo Gal, University of Haifa, Haifa, Israel

Based on the work of the Numeracy Expert Group for the Adult Literacy and Lifeskills (ALL) Study, which at the time consisted of:

Iddo Gal, University of Haifa, Haifa, Israel

Mieke van Groenestijn, HU University of Applied Sciences, Utrecht, Netherlands

Myrna Manly, Independent Numeracy Consultant, Reno, USA

Mary Jane Schmitt, TERC, Cambridge, Massachusetts, USA

Dave Tout, Australian Council for Educational Research, Melbourne, Australia

The other authors would like to acknowledge that sadly Myrna Manly died in November, 2012, and Mary Jane Schmitt died in January 2015. Both Myrna and Mary Jane made significant contributions to adult numeracy and mathematics education and research, not only across the USA, but also internationally.

Please cite the work as follows:

Tout, D., Gal, I., van Groenestijn, M., Manly, M., & Schmitt, M. J., (2020). *PIAAC Numeracy Task Complexity Schema: Factors that impact on item difficulty*. Australian Council for Educational Research. https://doi.org/10.37517/978-1-74286-609-3.

CONTENTS

1. ABSTRACT

This paper describes some lessons learned from international adult numeracy assessments that can help in understanding the challenges that people, including both adults and school students, have when solving numeracy tasks and their levels of performance on functional mathematical problems. The paper presents a theoretical schema of five factors that predict, separately and in interaction, the complexity or level of difficulty of mathematically-related assessment tasks, including tasks that incorporate texts and require literacy or reading skills, which are very common in adults' lives. The model was originally developed as part of the development of the Adult Literacy and LifeSkills survey in the mid '90s, but later adapted and effectively used within the Programme for the International Assessment of Adult Competencies (PIAAC), a.k.a. OECD Survey of Adult Skills.

The five "complexity factors" described in the model are grouped into two factors addressing mainly textual aspects of tasks, and three factors addressing the mathematical aspects of tasks. These factors can assist test developers, researchers and educators in predicting task difficulty and in targeting the development of items and tasks to more efficiently cover the range of student performance and skill levels.

2. INTRODUCTION

This paper describes a schema of five factors that can be used to explain as well as predict the difficulty of different numeracy assessment questions (items) and tasks. The schema was developed for the Adult Literacy and Lifeskills (ALL) survey by the ALL Numeracy Expert Group (Gal et al, 2005), and has also been used in the numeracy assessment for the Programme for the International Assessment of Adult Competencies (PIAAC) (PIAAC NEG, 2009; Tout et al., forthcoming). Although the complexity schema was developed within the context of designing specific international assessments, it has broad relevance for researchers, test developers and educators working in diverse contexts.

2.1. Preface

There are many challenges in writing quality test items in numeracy or mathematical literacy, since there are many factors and aspects to be taken into account (see Tout & Spithill, 2015). One key challenge faced by test developers for relatively large scale or high stakes assessments, which are aimed at a broad and diverse populations (of adults or school students), is the need to create a test with items that are suitable for estimating multiple proficiency levels within target populations. This means that the test needs to contain items that range from easy to difficult or challenging, in order to be able to scale students across the full range of the proficiency continuum.

The above means that test developers need to be able to estimate item parameters and anticipate each question's relative level of difficulty *in advance*, i.e., *before* any piloting and before the collection of empirical data or psychometric evidence about the actual performance of the items in the field. Predicting the level of difficulty of items at an early stage (before piloting) is essential so that the final test can have a fair and reasonable spread of items across the breadth of the expected levels of skills of the target population.

Knowledge about the factors that affect item complexity is essential for test developers (whether they are assessment designers, researchers, or teachers) and has many advantages. First, such knowledge enables test developers to produce items of varying difficulty levels, as it enables them to know in advance *what parameters or factors to manipulate and how to adjust item difficulty up or down*, i.e., make items easier or harder, by changing particular aspects or characteristics in the items. Second, such knowledge helps to *reduce unnecessary factors* that cause an item to be harder than it should (e.g., overload of text), and this can improve item reliability and validity. Finally, a schema of complexity factors helps in the *interpretation* of the resulting statistics about actual performance on each item, as it enables developers to understand what factors shape the observed distribution of responses, and how to explain resulting differences between the relative levels of performance of persons (adults or students) who had different scores on the assessment.

2.2. Background: International surveys of adult skills

In order to understand the origins of the schema of complexity factors described in this paper, and how it connects with the Programme for the International Assessment of Adult Competencies (PIAAC), we need to first provide some background about international or comparative surveys of adult skills. (A brief summary in this regard is provided in Kirsch & Lennon, 2017). PIAAC builds on earlier international adult literacy surveys, beginning with the International Adult Literacy Survey (IALS) (OECD & HRDC, 1997; OECD & Statistics Canada, 1995) in the 1990s. IALS evolved to become the Adult Literacy and Lifeskills (ALL) survey (OECD & Statistics Canada, 2005) in the mid-2000s, and was further expanded into the Programme for the International Assessment of Adult Competencies (PIAAC) (OECD 2013, 2016a, 2019), which is designed to run every 10 years. PIAAC's first cycle was undertaken in almost 40 different countries between 2008 and 2018; the next cycle is now under way, and the assessment is scheduled for implementation in 2022-2023.

PIAAC is a programme auspiced by the Organisation for Economic Co-operation and Development (OECD). It aims to measure the key cognitive and workplace skills needed for individuals to participate in society and for economies to prosper. The evidence from PIAAC aims to help countries better understand how education and training systems can nurture these skills. Educators, policy makers and labour economists can then use this information to develop economic, education and social policies that will continue to enhance the skills of adults.

Throughout IALS, ALL and now PIAAC, several competencies with mathematical or quantitative aspects have been the target of assessment, subsumed within several key constructs: quantitative literacy, document literacy, mathematical literacy and numeracy. Gal et al. (2020), Gal & Tout (2014) and Tout (2020) provide detailed discussions of these constructs and the differences between them. For the purposes of this paper, it is important to understand how the Numeracy Expert Groups have viewed and used the term *adult numeracy.* Fundamentally it refers to people's capability to use a range of mathematical and statistical knowledge and skills to solve problems in the real world for a purpose. Thus, to be considered numerate, it is expected that people will need to know some mathematics, and be able to apply that knowledge within a real-world context.

The current definition of numeracy in PIAAC cycle 2 (Tout, 2020; Tout et al., forthcoming) is:

> *Numeracy is accessing, using and reasoning critically with mathematical content, information and ideas represented in multiple ways in order to engage in and manage the mathematical demands of a range of situations in adult life.*

2.3. Understanding difficulty levels: the creation of the numeracy task complexity schema

The initial Numeracy Expert Group (NEG) for ALL, formed in the late 1990s, undertook the development of an assessment of numeracy skills, which built on, and extended, the Quantitative Literacy assessment component of IALS. This development is documented in a range of reports including Gal et al. (2005) and Tout (2020).

The ALL NEG first developed an agreed international conceptual framework for numeracy and its assessment. This included a description of what was being assessed and why and how, and the assessment construct which described characteristics of the stimuli and questions, their type and style, their content, etc. (see Gal et al., 2005). The NEG also was asked to develop a theoretical schema that would enable the prediction, in advance of an assessment actually taking place, of how difficult each numeracy question was and to validate the schema empirically.

The schema needed to consider and describe the various factors affecting numeracy task complexity and difficulty. The schema was used internally by the item development team and the NEG for various purposes, e.g., to inform item design, to evaluate items chosen for inclusion in the final assessment, and to inform the descriptions or interpretations attached to different performance levels on the assessments. This research paper describes the development of this numeracy task complexity schema. The development was a complex process as it had to integrate approaches developed in the context of both mathematics and literacy assessments, since adult numeracy relates to both domains.

2.4. Task and text complexity in reading

The numeracy task complexity schema was able to build on and learn from the work on task and text complexity in reading. In over 30 years of national and international surveys of adult skills, especially emanating from the International Adult Literacy Survey (IALS) and earlier studies, the components of task and text complexity and the variables that interact to determine the level of difficulty of reading tasks have been researched and schemas developed. Key research in this area include those of Kirsch and Mosenthal (1990); Kirsch, Jungeblut and Mosenthal (1998) and Kirsch (2001). This work has been instrumental in the understanding about teaching and learning of literacy skills. The basis of this has been directly attributable to the methodology behind international surveys, Item Response Theory (IRT), and the ability to conduct research using the empirical data from the surveys.

The research fundamentally argues that a number of variables interact to determine the level of difficulty of reading tasks. The variables relate to the structure and complexity of the text, to the nature of the task (i.e., the relationships between the text and the question being asked), and to the nature of the processes or strategies that relate the information in the question to information in the text. The work and development of the numeracy task complexity schema outlined in this paper derived much of its thinking and structure from this work.

2.5. Updates and refinements to the original ALL task complexity schema

The numeracy task complexity schema outlined in the following sections has been updated since the previous version was published in the PIAAC cycle 1 Numeracy framework document (PIAAC Numeracy Expert Group, 2009). The schema remained fundamentally the same from ALL through to PIAAC cycle 1.

The updates to the task complexity schema since then are based on two parallel uses of the schema. In numeracy test development work at the Australian Council *for* Educational Research (ACER), the schema was found to be highly useful in predicting item difficulty and in therefore targeting item writing more efficiently across the range of student performance and skill levels. The other use, which resulted in more significant enhancements to the schema, arose from when it was used as the basis for some comparative mapping work of different items from across different national and international mathematics and numeracy assessments by one of the authors and member of the current Numeracy Expert Group.

The updates and enhancements are mainly as a result of the existing schema being used and applied to numeracy (and mathematics) test items within a secondary school assessment context, where some test items had a more school based mathematics curriculum focus. This focus highlighted some gaps in the detailed descriptions sitting behind the schema (see the section, *Scoring for each of the Complexity Factors*). Whilst not seen as initially necessary for ALL and PIAAC, there was little specification included around some of the more formal, school based content and its associated terminology, representations and symbolism that can be used within the world of school mathematics. Hence some enhancements were made that highlighted this more explicitly within the two relevant existing factors that described the more mathematical aspects of task complexity, namely: *Factor 3. Complexity of mathematical information/data* and *Factor 4. Complexity of Type of operation/skill*. This included expanding on descriptions relating to symbols and conventions, algebra, the more formal aspects of the properties of shapes, along with an expanded specification about sense of number and estimating with numbers.

The updated schema was used by the NEG for PIAAC cycle 2 to estimate the difficulty level of the newly developed items, and to assist in the selection of the items for the field trial. The above enhancements were fortuitous and useful in this task for PIAAC cycle 2. The enhancements to the task complexity schema were relevant to two of the challenges and new endeavours for PIAAC cycle 2 test development. One was the development of the numeracy components which is attempting to target gathering more information about the skills of adults performing at the lower end of the proficiency scale. Number sense was the area that the NEG decided to attempt to assess, so the enhancements to the schema in relation to number sense were valuable in light of this endeavour. Similarly, the NEG were tasked with trying to develop new items up at the other end of the scale – at Level 5. Hence the elaborations about the more formal, mathematical terminology, representations and symbolism was a useful extension. These enhancements will also be useful when developing the updated numeracy proficiency descriptions for PIAAC cycle 2.

3. FACTORS AFFECTING THE COMPLEXITY OF NUMERACY ITEMS

The following text is mainly based on the version originally published in the ALL Technical Report (see Gal et al., 2005). This schema was found useful to inform item development, i.e., help in the creation of items that spread over a range of difficulty levels. Results from the ALL pilot study showed that predicted difficulty of items used by the schema described below was highly correlated with observed difficulty (r = 0.79). Because of the recursive nature of the testing of this schema (e.g., the same individuals wrote the schema and rated the complexity of items), caution should be exercised in further explicit, interpretive use of the present version. However, the schema's more recent use in other projects indicates it can be successfully used and applied. While further validation is needed, the schema in its current state nonetheless appears to be a useful tool for the development of test items and for the understanding and interpretation of testing results, and has been used for that purpose in a number of contexts.

3.1. Previous research on task complexity

In IALS, three factors were found to be the principal components of task difficulty (regarding literacy or text-based tasks): plausibility of distractors, type of match required, and type of information required. The difficulty of the Quantitative Literacy (QL) tasks appeared to be a function of several other factors:

1. The particular arithmetic operation required to complete the task
2. The number of operations needed to perform the task
3. The extent to which the numbers are embedded in printed materials
4. The extent to which an inference must be made to identify the type of operation to be performed (i.e. problem transparency; see below)

The IALS QL difficulty factors overall fit those used in large-scale assessments of mathematical skills (with children), which often make use of three or four factors:

1. The mathematical concepts involved: number systems and number sense, spatial and geometrical topics, functions and algebra, chance/statistics topics, etc. Concepts that are related to topics taught in lower grades are considered easier.
2. The complexity of operations: addition, subtraction, multiplication, and division, as well as dealing with whole numbers, with decimals, and with percents. Operations that are related to topics taught in lower grades are considered easier.
3. The number of operations: one-step problems are considered easier than multi-step problems.
4. Problem transparency: This factor is sometimes relevant; it refers to the extent to which the problem situation includes clearly identified numbers or entities and the extent to which it is clear what operations or actions to perform. To the extent that these are not clear or transparent, respondents have to extract needed information by applying comprehension and inference strategies, making the task more complex.

There are other adult-related assessment projects on which to draw to develop the levels of complexity. Both the Essential Skills Research Project in Canada and the Applied Numeracy sub-test of the Work Keys test battery (American College Testing, 1997) use a two-factor model of complexity in their description of numeracy levels. The first factor, "operations required;" is seemingly straightforward and refers to the difficulty of operations called for. However, this is complicated by the level of difficulty of the numbers being manipulated: computations that include fractions and decimals are usually more difficult than those with whole numbers.

The Essential Skills model spells out two sequences of complexity on this factor: *Operations* and *Translation* of information (sometimes called 'problem transparency').

Operations

1. Only the simplest operations are required and the operations to be used are clearly specified. Only one type of mathematical operation is used in the task.
2. Only relatively simple operations are required. The specific operations to be performed may not be clearly specified. Tasks involve one or two types of mathematical operation. Few steps of calculations are required.
3. Task may require a combination of operations or multiple-applications of a single operation. Several steps of calculation are required. (More advanced operations may call for multiplication or division.)
4. Tasks involve multiple steps of calculation.
5. Tasks involve multiple steps of calculation. Advanced mathematical techniques may be required (e.g., percents, ratios, proportions).

Translation (Problem Transparency)

1. Only minimal translation is required to turn the task into a mathematical operation. All the information required is provided.
2. Some translation may be required or the numbers needed for the solution may need to be collected from several sources. Simple formulae may be used.
3. Some translation is required but the problem is well defined.
4. Considerable translation is required.
5. Numbers needed for calculations may need to be derived or estimated; approximations may need to be created in cases of uncertainty and ambiguity. Complex formulae, equations or functions may be used.

Two considerations prompted us to question the appropriateness of using mathematics-related frameworks (from Essential Skills or elsewhere) as the sole source for development of a complexity schema for items assessing adults' ability to cope with real-world numeracy tasks. First, effective coping with many real-world quantitative problems depends upon people's ability to make sense of and interact with different types of texts. This is hardly recognised by the Essential Skills model. Hence, it was essential to add difficulty factors that acknowledge the inherent links between literacy and numeracy, quite similar to those used in IALS.

Another, albeit a more restricted consideration, is that the ordering of complexity of tasks by the type of operation performed may not be as clear with adults as it may be with children. Such ordering in school-based assessments is predicated on traditional school curricula, where more advanced topics are learned at higher grades. However, adults are known to use a lot of invented strategies, perhaps more so, and more efficiently so, than children. Multiplication or division problems, which can prove relatively hard for some young people, may be solved by seemingly simpler strategies, such as by repeated addition or repeated subtraction; complex numbers may be broken down in ways that ease mental load, and so forth. In addition, adults' familiarity with everyday contexts, such as with monetary entities, facilitates their performance with some seemingly advanced concepts. For example, specific benchmark values of fractions and percents, such as 1/2, 1/4, 50%, or 25%, are familiar to many people; as a result, they may be easier to manage than expected, violating curriculum-based ordering of difficulty. Hence, an overall complexity level has to be used, in order to weight these "inconsistencies" in ordering of difficulty levels proposed in other schemas.

3.2. The five Complexity factors

The above literature review suggests that a framework of factors affecting the complexity of numeracy tasks should not only address factors related to the numerical and textual aspects of tasks, but should also address other issues. It should treat separately the number of operations and the type of operations from the type of mathematical (or statistical) information to be processed, which may involve numbers explicitly but also other types of mathematical information. In so doing, the desired framework of complexity factors should take into account the broad scope of the definition of numeracy, i.e., reflect the variation within contexts, the range of mathematical ideas/content, the types of possible responses, and the types of representations that cut across adult life contexts.

With the above considerations in mind, five key factors have been identified that are predicted to affect, separately and in interaction, the difficulty level of numeracy tasks to be used in the ALL survey. These five "complexity factors" are outlined in Table 1 and are organised in two sets: two factors that address mainly textual aspects of tasks, and three factors that address the mathematical aspects of tasks. These five factors are listed separately for clarity of presentation, but in actuality are *not* independent of each other and do interact in complex ways. Each factor is examined in some detail below, followed by a later subsection that describes the calculation of an overall complexity level for each item, taking into account all five factors.

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

Table 1: Complexity Factors—Overview

1. Type of Match/Problem Transparency

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

In easy tasks, the type of information (e.g., numerical values) and the operations needed are apparent and obvious from the way the situation is organised. In more difficult ones, the values must be located or derived from other values; the operations needed may have to be discovered by the performer, depending on his or her interpretation of the context and of the kind of response expected. As well, numeracy situations may involve text to varying degrees, and this text may be of different degrees of importance. There may be a situation where there is little or no text. Some situations may involve pure quantitative information that is to be interpreted or acted upon with virtually no text or linguistic input. In other words, the performer derives all the information needed to respond from the objects present in the situation or from direct numerical displays.

At a higher level, some textual or verbal information may be present alongside the mathematical information. The text can provide background information about the problem situation, or some instructions. For example, a bus schedule, cooking instructions, and a typical school-type word problem all involve some text and some numbers. Still other situations would be heavily text-based or may not involve any numbers or mathematical symbols at all, just plain text. The task will contain mathematical or statistical information that a person needs to understand and, in some cases, act upon, but it will be much less transparent. It may be heavily embedded in dense text or may require using information from a number of sources within or even outside the text/task, or could also mean that outside information (e.g. the understanding and knowledge of a formal formula/process) may be needed to answer the question.

This factor requires that a task will be analysed in terms of the questions: How difficult is it to identify and decide what action to take? How many literacy skills are required? Is all the necessary information there?

2. Plausibility of distractors

This variable is literacy related, even though it can involve mathematical components. This concerns the extent to which information in the stimulus for the question shares one or more features with the information requested in the question but does not fully satisfy what has been requested. Questions are considered to be easiest when no distractor information is present in the material at all—everything that is needed to answer the question is there, it is explicit with no other distracting information available. Questions tend to become more difficult as the number of distractors increases, as the distractors share more features with the expected response. At higher levels of difficulty, tasks can involve irrelevant information both within the question as well as within the text. In terms of mathematical information, a low level of plausible distractors would mean that no other mathematical information was present apart from that requested, making the numbers or data required to use easy to identify. At a higher level, there may be either some other mathematical information in the task (or its text) that could be a distractor, or the mathematical information given or requested could occur in more than one place. For example, when the numbers required to undertake an arithmetic operation must be extracted from material that contains a range of similar, but irrelevant, information, the task becomes increasingly difficult.

This factor requires that a task will be analysed in terms of the questions: *How many other pieces of mathematical information are present?*

3. Complexity of Mathematical Information/data

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

This factor requires that a task will be analysed in terms of the question: How complex is the mathematical information that needs to be manipulated or managed?

4. Type of Operation/Skill

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

This factor requires that a task will be analysed in terms of the question: *How complex is the mathematical action that is required?*

5. Expected Number of Operations/Processes.

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

This factor requires that a task will be analysed in terms of the question: *How many steps and types of steps/processes are required?*

3.3. Overall Complexity Level

It is possible to estimate the overall difficulty level of a specific item by first scoring the item on each of the five factors of complexity, according to the levels described in the following section, and then summing together the scores for each factor. Figure 1 on the following page explains the process; the next section, *Scoring for each of the Complexity Factors*, details and provides scores for each level of the five factors in detail. The total summary score can range between 5 (easiest) and 19 (most difficult).

The estimation process outlined in Figure 1 suggests that each factor has a separate contribution to an item's overall difficulty or complexity. However, it can be hypothesised that as tasks become more complex, actual performance on items may increasingly depend not only on each factor by itself, but also on the interplay or interaction between them. Hence, the computational process suggested in Figure 1 can provide only approximate information about an item's anticipated difficulty level.

Further, the difficulty of a task cannot in some cases be predicted without taking into account characteristics of the person who interacts with the task. The same task may be more difficult for some individuals and less difficult for other individuals, depending on factors such as their familiarity with the context in which a task is situated, knowledge of formal mathematical notations, background world knowledge, as well as general literacy, problem-solving, and reasoning skills. For example, it could be predicted that a task that involves the composition of a fertilizer would be more difficult for an urban apartment dweller than for a rural farmer whereas a task that uses a bus schedule would be more difficult for the farmer. For the above reasons, the prediction of the difficulty of a task in isolation of detailed knowledge about the respondent himself can only be an estimate.

4. COMPLEXITY FLOW CHART

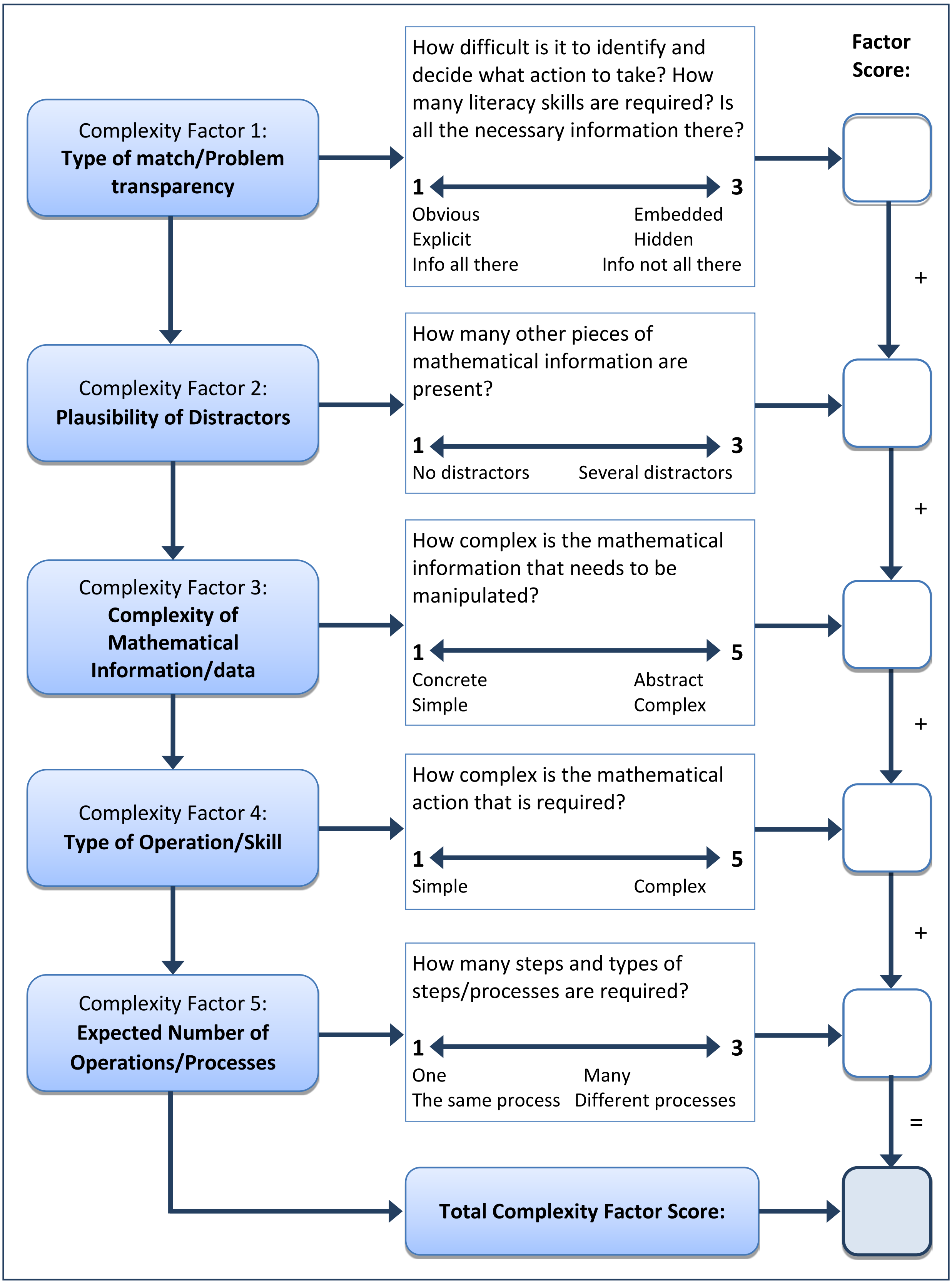


Figure 1. Complexity Flow Chart

5. SCORING FOR EACH OF THE COMPLEXITY FACTORS

Complexity Factor 1. Type of match/Problem transparency How difficult is it to identify and decide what action to take? How many literacy skills are required? Is all the necessary information there?		
Score 1	**Score 2**	**Score 3**
In the question and the stimulus, the information, activity or operation required: - is clearly apparent and explicit—and all required information is provided and where minimal translation or interpretation is required - is specified in little or no text, using simple, familiar and non-formal language/symbols, familiar objects and/or photographs or other clear, simple visualizations - is about locating obvious information or relationships only - closed question—not open-ended	In the question and the stimulus, the information, activity or operation required: - is given using clear, simple sentences and representations including some formal language/symbols and/or visualizations where some translation or interpretation is required - is located within a number of sources within the text/activity. - may need to bring to the problem simple information or knowledge from outside the problem. - fairly closed question	In the question and the stimulus, the information, activity or operation required: - is embedded in text including more technical or formal language/representations where considerable translation or interpretation is required and/or - may need to be derived or estimated from a number of sources within or outside the text/activity and/or - the information or action required is not explicit or specified or necessary information or knowledge is missing, so outside information or knowledge needs to be brought in - more complex, open-ended task

Complexity Factor 2. Plausibility of distractors How many other pieces of mathematical information are present?		
Score 1	**Score 2**	**Score 3**
- no other mathematical information is present apart from that requested—no distractors	- there is some other mathematical information in the task that could be a distractor - the mathematical information given or requested can occur in more than one place	- a range of other irrelevant mathematical information appears - mathematical information given or requested appears in several places.

Complexity Factor 5. Expected number of operations/processes How many steps and types of steps/processes are required?		
score 1	**score 2**	**score 3**
- one operation, action or process	- application of two or three steps, the same or similar operation, action or process Note: repeating the same sequence of operations/processes only counts once.	- integration of several steps covering more than one different operation, action or process

Complexity Factor 3. Complexity of mathematical information/data: How complex is the mathematical information that needs to be manipulated?				
score 1	**score 2**	**score 3**	**score 4**	**score 5**
Context Based on very concrete, real life activities, familiar to most in daily life.	Based on common, real life activities.	Based on real life activities, but less often encountered.	Based on real life activities but unfamiliar to most	Based on abstract ideas or unfamiliar activity in a context new to most.
Symbols and conventions - simple and informal symbolism, diagrams and conventions relevant to the mathematical knowledge of the level, e.g., 57, $5.98, ½ , +, –, x, ÷, =	– a combination of mainly informal and some formal symbolism, diagrams, graphs and conventions relevant to the mathematical knowledge of the level, e.g. %, 0.25, mL, °C/F, “”/cm, ( )	- a combination of both formal and informal symbolism, diagrams, graphs and conventions relevant to the mathematical knowledge of the level, e.g. 12.5%, km/hr, $/kg, <, >, ≤, ≥, ≠, ^{2}, ^{3}	- a combination of informal but mostly formal mathematical symbolism, diagrams, graphs, algebraic representation and conventions relevant to the mathematical knowledge of the level, e.g. A = 2πr; √2, -5°C	- a combination of specialised formal and general mathematical symbolism, diagrams, algebraic representation, graphs and conventions relevant to the mathematical knowledge of the level, e.g. Sin60°= √3/2, Σ
Quantity Whole numbers to 1,000 Fractions, decimals, percents: - benchmark fractions ($^1/_2$, $^1/_4$, $^3/_4$) - decimal fraction for a half only (0.5) and equivalent as a percentage (50%)	- large whole numbers including millions - other benchmark fractions, like $^1/_3$ and $^1/_{10}$ - common decimals, like 0.1, 0.25 to 2 decimal places - common whole number percents, like 25% and 10%.	- large whole numbers including billions - other fractions - decimals to 3 decimal places (other than money) - other percents - mixed numbers	- negative integers - recurring decimals	- all remaining types of rational (and some irrational) numbers including directed/signed numbers
Pattern and relationship/algebra - very simple whole number relations and patterns	- simple whole number rates and ratios - whole number relations and patterns	- rates and ratios - relations and patterns including written everyday generalizations/formulae (e.g., area/volume)	- complex ratios, relations, patterns - simple formula and algebraic expressions including inequalities	- formal mathematical information and expressions such as more complex algebraic expressions, formulae, knowledge of relationships between dimensions or variables, etc.
Measures/ Dimension/Space - standard monetary values - common everyday measures for length (whole units) - time (dates, hours, minutes) - simple, common 2D shapes - simple localised maps or plans (no scales)	- everyday standard measures for length, weight, volume , including common fraction and decimal units - common 3D shapes and their representation via diagrams, nets or photos - common types of maps or plans with visual scale indicators	- other everyday measures (area included) including fraction and decimal values - more complex 2D and 3D shapes, or a combination of 2 shapes, and their representation via diagrams, nets, incl geometric properties - area and volume formulae - common types of maps or plans with ratio type scales	- all kinds of measurement scales - complex shapes or combinations of shapes	
Chance/Data - simple graphs, tables, charts with few parameters and whole number values - simple whole number data or statistical information in text	- graphs, tables, charts with common data including whole number percents—whole number scales in 1s, 2s, 5s or 10s - data or statistical information including whole number percents	- graphs, tables, charts with more complex data (not grouped data) - more complex data or statistical information including common average, chance and probability values - scales: more complex whole number, fractional or decimal	- complex graphs, tables or charts including grouped data - complex data or statistical information including probabilities, measures of central tendency and spread	

Complexity Factor 4. Complexity of Type of operation/skill: How complex is the mathematical action that is required?				
score 1	score 2	score 3	score 4	score 5
<u>Communicate/Reason</u> - no explanation - a single simple response (orally, or in writing)	- a simple response required (orally, or in writing)	- simple explanation of a (level 1 or 2) mathematical process required (orally, or in writing) -	- explanation of a (level 3) mathematical process required (orally, or in writing)	- complex, abstract and generative reasoning or explanation required
<u>Compute</u> - a simple arithmetical operation (+, -, x, ÷) with whole numbers or money	- simple arithmetical operations (+, -, x, ÷) with decimals - calculating common fraction, decimal fraction and percentages of values - using common rates (e.g. $/lb.); time calculations; etc. - changing between common equivalent fraction, decimal and percent values, including for measurements e.g. $^1/_4$ kg = 0.250kg	- more complex applications of the normal arithmetical operations such as calculating with fractions and more complex rates, ratios, decimals, percentages, or variables - squares, cubes - simple probability calculations	- applications of other mathematical operations such as square roots' powers/exponents etc.	- more advanced mathematical techniques and skills e.g. trigonometry
<u>Sense of number and estimation</u> - counting, naming, comparing and place value understanding of whole numbers up to 1000 - Understanding the operations of (+, -, x, ÷) and their interrelationships	- naming, comparing and place value understanding of whole numbers up to millions - naming, comparing, understanding and equivalence of common fractions and percentages - estimating and rounding off, when requested, to whole number values or monetary units	- naming, comparing and place value understanding of all whole numbers and decimals - naming, comparing , and understanding of all fractions incl. equivalence of fractions and percentages - estimating and rounding off to requested number of decimal places	- making a contextual judgment re whether a found answer is realistic or not and changing the answer to the appropriate correct rounded (but not necessarily mathematically correct) answer.	
<u>Use formula/ model</u>	- evaluating a given simple formula involving common operations (+, -, x, ÷) expressed in real world terms/language	- using and solving simple, common formula and equations - generating, graphing and interpreting simple, common algebraic graphs	- developing/creating and using straight forward formulae - using strategies such as working backwards or backtracking (e.g. 15% of ? = $255) - using and solving simple inequalities - generating, graphing and interpreting more formal graphs	- generating, transposing & graphing more complex equations and formulae - using and interpreting standard formal algebraic and graphical conventions and techniques
<u>Measure/Shape properties</u> - knowing common straight forward measures and personal measures - naming, comparing common 2D shapes - comparing whole unit measurements	- visualizing/representing, comparing and describing 2D and 3D shapes, objects or geometric patterns or relationships, incl simple nets - estimating, making and interpreting standard measurements using common measuring instruments and scales	- using angle properties and symmetry to describe shapes or objects - transposing shapes (rotations/reflections) - understanding relationship between length/area - estimating, making and interpreting non-standard measurements - converting between standard measurement units within the same system - interpolating values on scales	- understanding more formal geometric representations and relationships e.g., parallel lines and angle relationships/properties - understanding relationships between area/volume - converting between non-standard measurement units within the same system	- converting between measurements across different systems
<u>Interpret</u> - locating/identifying data in texts, graphs and tables - orientating oneself to maps and directions such as right, left, etc.	- reading and interpreting data from texts, graphs and tables - following or giving straight forward directions	- generating, organising, graphing non-grouped data - interpolating data on graphs - calculating distances from scales on maps	- calculating common measures of central tendency & spread for non-grouped data - calculating permutations/combinations - extrapolating data - reading and interpreting trends and patterns on graphs, including slope/gradient	- graphing grouped data - calculating measures of central tendency & spread for grouped data

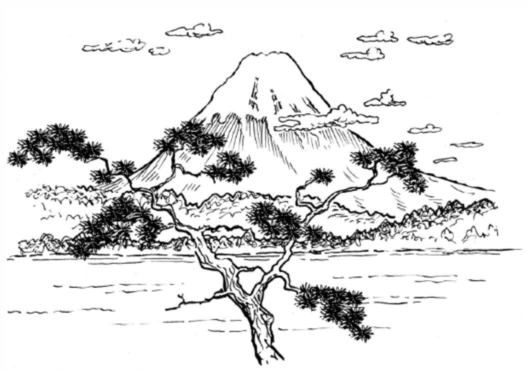
6. SAMPLE ITEM RATING

Below is an example of how the Complexity schema can be applied to an assessment item to estimate its relative difficulty.

The item below is one question from a publicly released mathematical literacy unit taken from the OECD's Programme for International Student Assessment (PISA). It was a paper-based unit from the 2012 survey administration. There were three questions based around the same context of Mount Fuji.

CLIMBING MOUNT FUJI

Mount Fuji is a famous dormant volcano in Japan.



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?

....

This question requires the respondent to interpret the context, the terms used and the different pieces of information and undertake a multi-step strategy to formulate and use a model (formula) that connects time, speed, and distance, and to implement calculations with time, including to work backwards to determine the required starting time.

In relation to using the numeracy complexity schema, you need to consider each factor independently and the role each takes in being able to solve the problem, and estimate how to score it against the more detailed descriptions in the above tables. In many cases more than one of the descriptions and related scores for Complexity Factors 3 (*Complexity of mathematical information/data*) and Factor 4 (*Complexity of Type of operation/skill*) will apply as problems may well require the understanding of different levels of mathematical information/data or the use and application of different skills or operations. Generally, the score given will be based on the highest level score from across the different elements within each factor, as this indicates the highest level cognitive demand of the problem solving process.

For this question, below are possible scores that could be given for each complexity factor.

1. Type of match/problem transparency (3/3)
    - This question is embedded in text where there is the need to translate and interpret what is written in order to be able to decide what to do – although it's not a long or complex text – so it could be a score of 2 or 3. However, because you need to bring to the problem knowledge from outside (the mathematical relationship between speed, distance and time), this probably makes a score of 3 more appropriate.

2. Plausibility of distractors (2/3)
    - There is the need to select and use the relevant and correct mathematical information from different sources embedded throughout the task (times, speeds, distances, etc.). A score of 2 is appropriate here as it doesn't really meet the requirements for a score of 3.

3. Complexity of mathematical information/data (4/5)
    - Understanding the mathematical relationship (formula) that connects speed, distance and time is critical here and being able to interpret it, pushes this up to a score of 4. Other information needed to solve the problem are at lower score points, but this is the critical piece of information needed to answer the question.

4. Complexity of Type of operation/skill (4/5)
    - There are a number of different operations and skills required here – the most challenging will be the application of the formula that connects speed, distance and time, which in this case requires working backwards to find the time taken given the speed and the distance. Hence this probably pushes it up from a score of 3 to closer to a 4.

5. Expected number of operations (3/3)
    - For this question there is clearly the need to integrate several steps covering more than one different process. So a score of 3.

So a possible total score here would be 16 out of a maximum possible score of 19, indicating this would be estimated by the schema to be a relatively difficult item on the PIAAC scale.

We do know from the results of PISA 2012 that this item was at the more difficult end for 15 year old students around the world. The data showed that across all countries only about 14% of students could answer this question correctly. It was a Level 5 item on the PISA proficiency scale, which has a highest level of level 6. So a score of 16/19 (or 15/19) appears appropriate.

7. DISCUSSION

As has been demonstrated during the work on the ALL and PIAAC assessments, the numeracy complexity schema described in this paper can be used to guide and better target the writing and development of numeracy assessment items. Specifically, the schema can be used to help rate, compare and identify the spread of item difficulty of the range of test items, prior to the delivery or trialling of the test items. The descriptions detailed in the Scoring for each of the Complexity Factors tables earlier can also be used inform the descriptions or interpretations attached to different performance levels on assessments.

However, a number of insights emanating from the numeracy task complexity schema are also useful for educators and researchers, including in relation to the teaching of numeracy. The complexity schema described shows that text-related factors play a major role in affecting the difficulty of numeracy tasks. This is also supported by findings from prior research on factors affecting complexity of reading tasks. This has implications for educators and teachers including, for example, that a numeracy teacher is also a teacher of literacy and language, and that teachers need to explicitly teach students how to 'read' or 'excavate' the mathematical content embedded in texts within numeracy tasks. The other factors in the complexity schema show that in teaching numeracy there is the need to explicitly address a range of cognitive operations and content areas, including the complexity of the mathematical information; the type of operation/skill, and the impact and complexity of the number and types of operations or processes involved in solving the task at hand.

7.1. Further steps

Whilst as mentioned earlier there is some empirical support for the usefulness and application of the complexity schema, it needs to be noted though that its application is obviously a subjective process based on professional judgement. The process and results of using a complexity schema are best moderated across different reviewers to come to agreed understandings, positions and perspectives of the different factors and their detailed descriptions. This was the process used by the original NEG development team, who scored the items independently, then moderated their scores collectively, arguing and negotiating to an agreed score. We note that similar 'professional judgment' demands exist in many other aspects of developing and fine-tuning other types of assessment scales.

This moderation process is important in applying such a schema for rating or mapping purposes. Training teachers or test developers in the schema is important prior to using the item-rating process. The experience gained by the NEG points to the critical importance of having new judges rate sample questions individually and then moderate the results together as a team. This is critical for the successful use of the schema and for any calibrations based on the judgments of users of the complexity schema.

Further research could be undertaken to shed more light on the validity and usefulness of the numeracy task complexity schema. Formal studies could include research based on independent use and application of the schema, backed by empirical data analysis. This would potentially lead to further refinements and improvements to the current schema.

7.2. Conclusion

Despite some limitations as outlined earlier, the schema of complexity factors developed for numeracy assessment originally in ALL, but now updated and used in PIAAC Cycle 2, has made an important theoretical contribution to understanding the factors affecting difficulty levels of different numeracy tasks and questions. It provides a conceptual basis for predicting the different levels of complexity of a broad range of items well beyond those involving arithmetic operations only.

The *Numeracy Task Complexity Schema* described in this paper can be useful in multiple ways and facilitate the work of test development teams, since it:

- describes the different parameters that impact on item difficulty for numeracy tasks and test items
- helps to efficiently develop an appropriately targeted set of numeracy test items to meet the expected performance of the cohort being assessed
- enables contrasting and comparing different numeracy tasks or test items in terms of their relative difficulty or rating on five common underlying factors
- assists in the description of different levels of performance on numeracy tasks and test items
- informs numeracy and mathematics educators about different factors that need to be addressed in the teaching and learning of numeracy and mathematical literacy.

8. REFERENCES

American College Testing. (1997). *Work keys: Preliminary technical handbook*. IA: Author.

Gal, I., & Tout, D. (2014). *Comparison of PIAAC and PISA Frameworks for Numeracy and Mathematical Literacy*. OECD Education Working Papers, No. 102. OECD Publishing.

Gal, I., Grotlüschen, A., Tout, D., *et al.* (2020). Numeracy, adult education, and vulnerable adults: a critical view of a neglected field. *ZDM Mathematics Education,* 52**,** 377–394. https://doi.org/10.1007/s11858-020-01155-9.

Gal, I., van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2005). Adult numeracy and its assessment in the ALL survey: A conceptual framework and pilot results. In Murray, S. T., Clermont, Y., & Binkley, M. (Eds.). *Measuring adult literacy and life skills: New frameworks for assessment* (pp. 137-191). Statistics Canada. https://www150.statcan.gc.ca/n1/en/pub/89-552-m/89-552-m2005013-eng.pdf.

Kirsch, I. S. (2001). The framework used in developing and interpreting the International Adult Literacy Survey (IALS). *European Journal of Psychology of Education,* 16, 335–361. https://doi.org/10.1007/BF03173187.

Kirsch, I. S., & Lennon, M. L. (2017). PIAAC: a new design for a new era. *Large-scale assessments in education*, *5*(1), 11. https://link.springer.com/article/10.1186/s40536-017-0046-6.

Kirsch, I. S., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*(1), 5–30. https://doi.org/10.2307/747985

Kirsch, I. S., Jungblut, A., & Mosenthal, P. B. (1998). The measurement of adult Literacy. In S. T. Murray, I. S. Kirsch, & L. B. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey* (pp. 105-134). National Center for Education Statistics, U.S. Department of Education.

OECD & Human Resources Development Canada (HRDC). (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey*. OECD Publishing. https://vital.voced.edu.au/vital/access/services/Download/ngv:32828/SOURCE2?view=true

OECD & Statistics Canada (1995). *Literacy, economy and society: Results of the first International Adult Literacy Survey.* OECD Publishing/Statistics Canada. https://www.oecd.org/education/skills-beyond-school/41529765.pdf

OECD & Statistics Canada. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey.* OECD Publishing. https://doi.org/10.1787/9789264010390-en.

OECD & Statistics Canada. (2011). Literacy for Life: Further Results from the Adult Literacy and Life Skills Survey. OECD Publishing. https://doi.org/10.1787/9789264091269-en.

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

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

OECD. (2019). *Skills matter: Additional results from the Survey of Adult Skills*. OECD Skills Studies series. OECD Publishing. https://doi.org/10.1787/1f029d8f-en.

PIAAC Numeracy Expert Group (NEG). (2009). *PIAAC Numeracy: A Conceptual Framework.* OECD Education Working Papers, No. 35. OECD Publishing. https://doi.org/10.1787/220337421165.

Pimentel, S. (2013). *College and Career Readiness Standards for Adult Education.* U.S. Department of Education (DoE) Office of Vocational and Adult Education. http://lincs.ed.gov/publications/pdf/CCRStandardsAdultEd.pdf.

Tout, D. (2020) Evolution of adult numeracy from quantitative literacy to numeracy: Lessons learned from international assessments. *International Review of Education, 66,* 183-209. https://doi.org/10.1007/s11159-020-09831-4.

Tout, D., & Spithill, J. (2015). The challenges and complexities of writing items to test mathematical literacy. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy: the PISA experience* (pp. 145–171). Springer. https://doi.org/10.1007/978-3-319-10121-7_7.

Tout, D., Demonty, I., Díez-Palomar, J., Geiger, V., Hoogland, K., & Maguire, T. (forthcoming). *PIAAC numeracy: An assessment framework for the second cycle of the Programme for the International Assessment of Adult Competencies*. OECD Publishing.