# Lessons learnt from international assessments: Part 2

*By Dave Tout*

**In Part 2 of this two part series Dave Tout emphasises an often overlooked value of international assessments–they provide an insight and description of task development and complexity that teachers can refer to when creating their own classroom tasks and assessments–see Chris Tully's article in Numeracy Matters of this issue, 'The complexity scheme', for an example. Part 1 of Dave's article appeared in *Fine Print* vol: 36 no. 2, our previous issue.**

## Introduction

This is the second article about what can be learnt from international adult literacy and numeracy assessments—this time more related to teaching and learning. As described in the first article, when I began working on these assessments I was ignorant about the complex processes, sophistication and the theoretical frameworks sitting behind international assessments. There is information relevant to teaching and learning that stem from both the theoretical frameworks, constructs and descriptions of the assessments and also from the research that occurs based on the rich data set of empirical information about adults' literacy and numeracy performance, including in relation to the background data. My focus here will mainly be on numeracy as that is my key area of interest and expertise, but I will also include snippets about literacy and hope that the references provided will allow interested readers to follow up the available research papers and reports.

It seems to me that reading behind and beyond the initial and media focussed messages from ALLS and PIAAC about the results to look at the frameworks and what the related research tells us about teaching and learning, is equally important as the results themselves. Building on the empirical and theoretical research emanating from such international assessments and their frameworks strengthens the links between testing, research, and practice. Professionally, being involved in this research and development has enhanced my knowledge, and I have used it in my work on a range of projects, including training and the writing of curriculum and assessment frameworks such as the ACSF and the CGEA.

## Literacy frameworks, descriptions and research

As mentioned in the first article, sitting behind each survey is a theoretical framework of literacy and numeracy that attempts to describe what literacy and numeracy in the 21st century incorporates. Each of the assessments involves the development and publication of comprehensive reports and analyses of each of the domains being assessed.

The Adult Literacy and Life Skills Survey (ALLS) developed out of earlier national and international assessments including the International Adult Literacy Survey (IALS) (Statistics Canada and OECD, 1996; 1997). The Adult Literacy and Life Skills Survey (ALLS) (Statistics Canada and OECD, 2005; OECD Statistics and Canada, 2011) described knowledge and skills in prose literacy, document literacy and numeracy. In the follow up to IALS and ALLS, the Program for the International Assessment of Adult Competencies (PIAAC) (Jones et al, 2009), literacy was described as understanding, evaluating, using, and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential. The PIAAC literacy framework elaborated a range of characteristics including:

- medium—pen and paper, digital
- text type—argumentation, description, exposition, instruction, narration, records
- social contexts—work, personal, community, education
- task aspects—access and identify, integrate and interpret, evaluate and reflect.

## 'Reading components'

For the first time PIACC incorporated a lower level assessment of reading, for adults who are identified by a screener test to have low levels of literacy. This aspect of PIAAC is called the 'Reading components' assessment. In previous surveys, the information on the reading abilities of adults with poor skills was often insufficient to get an understanding of their difficulties and build a description of their abilities. The assessment of component skills is intended to provide a greater level of information about the skills of individuals with low levels of literacy proficiency. The reading components assessment assesses:

- word meaning of everyday words

- sentence processing—making a judgement of whether the sentence is sensible
- passage comprehension—time taken to read passage and choosing between correct and incorrect words to gain meaning.

Although the data and results from the PIAAC 'Reading components' are yet to be analysed for Australian adults, the international research and data that will be generated will add further to our knowledge about the basis and the beginnings of literacy development in adults. As such the assessment frameworks and the material sitting alongside them offer valuable information and background to literacy development for adults and provide a synthesis of research about the assessment and description of reading skills. They provide insights into what can be taught and what components of teaching reading are crucial for successful development of literacy skills. These can also be utilised in the development of curriculum and assessment frameworks.

## Benefits of item response theory: task and text complexity

In over thirty years of national and international surveys of adult skills, especially emanating from the 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 schemes developed. Key research in this area include those of Kirsch and Mosenthal, 1990; Kirsch and Mosenthal, 1994; 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.

As briefly described in the first article, IRT is the statistical methodology that sits behind international comparative surveys such as ALLS and PIAAC. IRT allows a large number of items of varying difficulty to be developed to assess a wide range of skills across a domain, and the test items can be placed on a scale of difficulty relative to each other, independent of the ability of students taking the test. By being able to align both the performance of people and the difficulty of tasks along such a common scale allows an analysis of the tasks with similar score values—do they share certain characteristics, how do they differ from one end of the scale to the other? This research enables a deeper understanding of the range of variables that contribute to item and task difficulty.

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.

While a wide range of characteristics have been described and analysed for each of these variables, there are three variables worth noting here that have been found to be highly related to the difficulty of a reading task. These were type of match, plausibility of distractors, and type of information. These are described further below.

Type of match relates to the relationship between what the learner or reader is asked to do in the question and the presentation of the required information in the text. Matching strategies are described including locating, cycling, integrating, and generating. For example, does the reader merely need to locate the information, or do they need to cycle through the text to identify two or more pieces of conditional information, or compare or contrast pieces of information?

Plausibility of distractors concerns the extent to which information in the text shares one or more features with the information requested in the question, but does not fully satisfy what has been requested. Tasks are judged to be easiest when no distractor information is present in the text. They tend to become more difficult as the number of distractors increases, as the distractors share more features with the correct response, and as the distractors appear in closer proximity to the correct response.

Type of information relates to the concreteness or abstractness of the requested information. The easier tasks are judged to be when the information is the most concrete whilst the difficult tasks to process were those with more abstract information.

Further details of all these factors and the research based on them, including their descriptions of levels of difficulty and schemes for estimating task and item difficulty for reading different types of texts are elaborated in a number of the above reports.

Similar and related research about text and task complexity has been undertaken in relation to the literacy assessment of fifteen-year-olds as part of the Programme for International Student Assessment (PISA). Not surprisingly there are many similarities, but also differences, between the schemes for fifteen-year-olds and those for adults. For example, see Lumley et al (2012) for the elaboration and research about ten different variables impacting on reading difficulty based on PISA.

Along with the descriptions of literacy described in the frameworks, and the different text and task complexity factors, a number of lessons can be taken that teachers and trainers can use in their literacy teaching. Examples include the need to ensure teaching tasks and activities cover a range of social contexts and text types; teaching students how to *read* the task or question; helping students develop strategies to access and identify information, to integrate and interpret (relate parts of texts to each other); and evaluating and reflecting (drawing on knowledge, ideas and values external to the text).

## Numeracy frameworks, descriptions and research

In ALLS, numeracy was defined as the knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations. In PIAAC, the follow up survey to ALLS, numeracy was described as the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life. This was elaborated to describe more detailed components of numeracy as shown in Table 1.

Similarly to literacy, this description and elaboration of numeracy can serve to highlight the complexity and extent of numeracy, and the range of factors that impact on someone becoming numerate. This can be utilised to help describe curriculum and assessment frameworks, alongside providing guidance about what elements need to be addressed and covered in numeracy teaching and learning to assist adults to develop a range of numeracy skills.

## Task complexity in numeracy

As mentioned previously, a significant challenge, and an unknown, was that the expert group for any new domain in international assessments had to develop a scheme describing the various factors affecting item complexity and difficulty. This theoretical construct had to be capable of predicting, in advance of an assessment actually taking place, how difficult each item was going to be and that this needed to be validated empirically. These schemes are used internally by item development teams and expert groups 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. Further description of this aspect of the assessments follows.

**Table 1. PIAAC's specification of numerate behaviour and its facets, From PIAAC numeracy expert group, 2009: 21–22**

*Numerate behaviour involves managing a situation or solving a problem:*

1. in a real context
   - everyday life
   - work
   - societal
   - further learning
2. by responding
   - identify, locate or access
   - act upon, use: order, count, estimate, compute, measure, model
   - interpret
   - evaluate / analyse
   - communicate
3. to mathematical content/ information/ ideas
   - quantity & number
   - dimension & shape
   - pattern, relationships, change
   - data & chance
4. represented in multiple ways
   - objects & pictures
   - numbers & mathematical symbols
   - formulae
   - diagrams & maps, graphs, tables
   - texts
   - technology-based displays.

*Numerate behaviour is founded on the activation of several enabling factors and processes:*

- mathematical knowledge and conceptual understanding
- adaptive reasoning and mathematical problem-solving skills
- literacy skills
- beliefs & attitudes
- numeracy-related practices and experience
- context/world knowledge.

## The ALLS/PIAAC numeracy complexity scheme

In relation to numeracy, PIAAC (based on ALLS) developed a complexity scheme to predict the difficulty or complexity of a numeracy assessment task. This was empirically validated in the ALLS work. A unique scheme of five factors was researched and written that attempted to account for the difficulty of different tasks, enabling an explanation of observed performance in terms of underlying cognitive processes or factors. Table 2: A summary of PIAAC complexity factors summarises the five

**Table 2. A summary of PIAAC complexity factors, from PIAAC numeracy expert group, 2009: 21-22.**

| Aspects | Category | Range |
|---|---|---|
| Textual | *Type of match/problem transparency*<br>How difficult is it to identify and decide what action to take?<br>How many literacy skills are required? | Obvious/explicit to embedded/hidden |
| | *Plausibility of distracters*<br>How many other pieces of mathematical information are present?<br>Is all the necessary information there? | No distractors to several distractors |
| Mathematical | *Complexity of Mathematical information/data*<br>How complex is the mathematical information that needs to be manipulated? | Concrete/simple to abstract/complex |
| | *Type of operation/skill*<br>How complex is the mathematical action that is required? | Simple to complex |
| | *Expected number of operations*<br>How many steps and types of steps are required? | One to many |

factors, and shows that two of them relate to textual aspects of numeracy tasks, and three relate to mathematical aspects of tasks. For each of these factors a detailed description was developed against a scoring system in the range from one through to three or five, for a total difficulty score in the range five to nineteen. These five factors and associated scoring schema are described more fully in the Annex of the PIAAC numeracy framework (Gal et al., 2009).

As with text and task complexity research in relation to reading, a number of lessons can be taken that teachers and trainers can use in their numeracy teaching. This includes, for example, that a numeracy teacher is also a teacher of literacy and language, and that you need to teach students how to *read* the text, and the task or question—how to excavate the maths from the context. That in teaching numeracy you need to create tasks and teach explicitly the range of cognitive operations and content areas, including the complexity of the mathematical information, the type of operation/skill, and about the impact and complexity of the number and types of operations involved.

## Deepening our understanding

As an example of further potential for research sitting behind surveys such as ALLS and PIAAC, in the pilot for ALLS, the expert group described and developed sets of parallel items for three different types of percentage questions related to shopping:

> Type 1: working out a percentage of an amount—How much would you have to pay for a TV advertised at 15% off, where the original price was $300?
>
> Type 2: working out what percentage is saved given the original price and the discounted price—What percentage is saved on a TV advertised at $255, where the original price was $300?
>
> Type 3: working backwards to find the original price given the percentage discount and the sale price—What was the regular price on a TV advertised at $255, where the percentage discount was 15%?

There were two sets of three items where the team was able to directly compare respondent's success on the three different types of percentage items. One was the TV items using 15% and the other was using a watch at $90 with a discount of 50%. Table 3 shows the results (percentage correct) on the sets of items covering all three types.

It appears from these results, that when the context is familiar (a 50% off sale) and the numbers are simple (45 is 50% of 90), adults show that they can solve all three kinds of percent problems equally well. However, when the percent is no longer what we consider a *benchmark* percent and the numbers are more difficult, the performance of adults on the three types of percent problems vary according to expectations, that is, they do best on Type 1 problems and worst on Type 3 problems.

Along with this data, including from other research into the ALLS results, and in work on other assessments, a number of clear messages result about how and what adults can do and cannot do. In particular, these include that the order and hierarchy of number skills are not as per school type curriculum—that adult numeracy skills and knowledge is much broader and chunky. For example, as per the above research, that common percentages and fractions (such as half or 50%, and 10% or 1/10th) can be seen as on about the same level of difficulty as working on the

**Table 3**

| | Type 1 | Type 2 | Type 3 |
|---|---|---|---|
| Watch (50%) | 93.3% | 90.8% | 92.1% |
| TV (15%) | 81.7% | 72.4% | 40.2% |

four operations with whole numbers. Another example from such international and national assessments is that asking questions related to more formal aspects of school type mathematics (e.g. place value) is more difficult than equivalent questions phrased in more contextually based formats.

## Conclusion

As I said at the end of the first article, I believe that the results, the research and the underpinning conceptual frameworks for international assessments such as ALLS, PIAAC and PISA add to the expertise and knowledge of both education and research communities. The empirically based research emanating from such assessments, alongside associated theoretical works such as around text and task complexity for literacy and numeracy, can and do contribute to enhancing and supporting teaching and learning. Building on the empirical research from such international assessments and their frameworks strengthens the links between testing, research, and practice.

**Dave is an experienced numeracy educator who is interested in making mathematics relevant, especially for disengaged students. He has worked in programmes in schools, TAFEs, ACE providers, universities, AMES and industry. Over the last fifteen years he has been involved in the development of the numeracy components of the International Adult Literacy and Lifeskills Survey (ALLS) and the Programme in Assessment of Adult Competencies (PIAAC). Dave joined the Australian Council for Educational Research in 2008 where he is the manager of Vocational, Adult and Workplace Education Services.**

## References

Gal, I., Alatorre, S., Close, S., Evans, J., Johansen, L., Maguire, T., Manly, M. & Tout, D. (2009). *PIAAC numeracy: A conceptual framework. OECD Education Working Papers,* No. 35. OECD Publishing. http://dx.doi.org/10.1787/220337421165

Jones, S., Gabrielsen, E., Hagston, J., Linnakylä, P., Megherbi, H., Sabatini, J., Deutsches, Tröster, M. & Vidal-Abarca, E. (2009). *PIAAC literacy: A conceptual framework. OECD Education Working Papers, No. 34.* OECD Publishing. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=edu/wkp(2009)13&doclanguage=en

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

Kirsch, I.S., Jungeblut, A. & Mosenthal, P.B. (1998). The Measurement of Adult Literacy. In Murray, T.S., Kirsch, I.S. & Jenkins, L. (Eds). *Adult literacy in OECD countries: Technical report on the first international adult literacy survey.* Washington, DC: US Department of Education, National Centre for Education Statistics.

Kirsch, I.S. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured.* Princeton, NJ: Education Testing Service.

Kirsch, I.S. and Mosenthal, P.B. (1994). Interpreting the IEA Reading Literacy Scales. In M. Binkley, K. Rust, and M. Winglee (Eds.), *Methodological issues in comparative educational studies: The case of the IEA Reading Literacy Study.* Washington, DC: National Center for Education Statistics, United States Department of Education.

Lumley, Tom, Routitsky, Alla, Mendelovits, Juliette, and Ramalingam, Dara. (2012). *A framework for predicting item difficulty in reading tests.* http://research.acer.edu.au/pisa/5

Mosenthal, P.B. and Kirsch, I.S. (1998). A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. *Journal of Adolescent and Adult Literacy,* 41(8) pp. 638–657.

OECD & Statistics Canada. (2011). *Literacy for life: Further results from the Adult Literacy and Lifeskills Survey.* OECD Publishing. Online: http://dx.doi.org/9789264091269-en

Sabitini, J. P. and Bruce, K. M. (2009). *PIAAC Reading Components: A conceptual framework. OECD Education Working Papers,* No. 33. OECD Publishing.

Statistics Canada and OECD. (1996). *Literacy, economy, and society: First results from the International Adult Literacy Survey.* Ottawa: Author.

Statistics Canada and OECD. (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey.* Ottawa: Author.

Statistics Canada and OECD. (2005). *Learning a living: First results of the Adult Literacy and Lifeskills Survey.* Ottawa and Paris: Author. Online: http://www.oecd.org/dataoecd/44/7/34867438.pdf