JEFF EVANS

3. WHAT TO LOOK FOR IN PIAAC RESULTS

How to Read Reports from International Surveys

INTRODUCTION

In October 2013, results from the Organisation for Economic Co-operation and Development (OECD) sponsored Programme for the International Assessment of Adult Competencies (PIAAC) for 24 participating countries (mostly in Europe, but also including North America, the Far East, and Australia) became available. As the successor to the International Adult Literacy Survey (IALS) in the 1990s and the Adult Literacy and Lifeskills (ALL) survey in the 2000s, PIAAC aims to provide information as an international comparative survey. It also has many similarities with national studies, such as Skills for Life in the United Kingdom (UK). Unlike international school level surveys (e.g. Programme for International Student Assessment [PISA], Trends in International Mathematics and Science Study [TIMSS][1]), which gain access to 'captive populations' in schools, PIAAC has needed to use a combination of household survey and educational testing methodologies. It represents a development from the earlier studies, in several ways:

- its first round covers a greater number of countries (24, two thirds of which are European Union (EU) members) – though probably all could be called 'advanced industrial societies'
- it focuses on three domains or 'competencies' – Literacy, Numeracy, and now Problem-solving in technologically rich environments (PSTRE)
- it uses computer administration, which, amongst other things, allows 'adaptive routing', aiming to assess the broad 'skill level' of the respondent from a few initial responses, and then to administer more appropriate items (in terms of difficulty) throughout the interview
- it implements a number of methodological and fieldwork improvements, for example, specification and regulation of sampling and fieldwork standards, and
- it has made its data available more quickly and more conveniently.

In addition, PIAAC is designed to be repeated, in order to build up time series data for participating countries. This 'longitudinal' feature would aim to increase the possibility of evaluating competing causal explanations using the study over time of correlations of the outcomes with relevant social or attitudinal variables.

K. Yasukawa & S. Black (Eds.), Beyond Economic Interests, 41–58.

In this chapter, I focus on how to understand these studies, by considering conceptual issues, methodological aspects (research design and execution), and presentation of results. I also discuss the types of results from Australia made available in October 2013, as well as preliminary results released by the Australian Bureau of Statistics (ABS) earlier in that year (ABS, 2013a, 2013b). The chapter aims to air questions concerning the relevance of these survey results to literacy and numeracy researchers and practitioners, and the types of further research possibly needed, in different national and local contexts.

POLICY CONTEXT

Educational policy is currently being developed on a world-wide scale, with supra-national organisations being key agencies for change (Rizvi & Lingard, 2010). In this context, the idea of Lifelong Learning (LLL) is central to the conceptualisation and development of adult literacy and adult numeracy. In international policy debates, LLL has been much contested, e.g. between 'humanistic' and 'economistic' approaches (Evans, Wedege, & Yasukawa, 2013). In this connection, it is important to consider work done both within the UNESCO programmes (e.g. Guadalupe, 2015), and by the OECD.

Here I focus on the OECD, the sponsor of PIAAC. OECD's view of LLL aims to promote several objectives:

- development of knowledge and competencies enabling each citizen to actively participate in various spheres of globalised social and economic life
- a broad view of learning, to include more than just the acquisition of technical skills for the economy (OECD, 2007, pp. 9–10)
- emphasis on the citizen's need to acquire *and update* a range of abilities, attitudes, knowledge and qualifications over the life-course, and hence the individual learner's responsibility for their own education (e.g. Walker, 2009)
- change in the focus of learning 'from what people know' to 'what they can do' (Moore & Jones, 2007), and
- weakening of the distinction between formal and informal education (Young, 2010).

Some of the consequences of these positions will be discussed below (see also Tsatsaroni & Evans, 2013).

The European Union (EU) is working closely with the OECD on PIAAC. Increasing globalisation and competitive economic environments are leading national governments to seek competitive advantage, "frequently defined in terms of the quality of national education and training systems judged according to international standards" (Brown, Halsey, Lauder, & Wells, 1997, pp. 7–8). Results from surveys like PIAAC (and PISA) may provide relevant international yardsticks.

For supra-national institutions like the EU, the area of LLL provides a domain where they can make a legitimate policy intervention, since, in a 'globalised' world,

a focus on labour mobility makes LLL a supra-national concern. This provides a basis for OECD's and EU's actions, leading to the promotion of the 'skills and competencies agenda', in all sectors of education and training (Grek, 2010). More generally, the OECD and the EU are disseminating ideas and practices that strongly influence national policy making around the world. These include:

- the promotion of expertise in creating comparable datasets, so that countries can measure the *relative* success of their education systems and shift policy orientations accordingly
- new forms of 'soft governance' of national educational systems, encompassing the production and dissemination of knowledge, and of comparative data such as educational and social indicators, and peer reviews involving country and thematic reviews – so that these supra-national organisations are 'governing by data' (Ozga, 2009).

Thus, one of the effects of international studies like PISA and PIAAC is to contribute to a 'comparative turn' in educational policy-making and to a "scientific approach" to political decision-making (Grek, 2010, p. 398).

THE PIAAC SURVEY

PIAAC's wider objectives were presented by Andreas Schleicher (2008) of the Education Directorate at OECD – as helping the participating countries to:

- *Identify and measure differences* between individuals and across countries in key "competencies"
- *Relate measures of skills* based on these competencies to a range of *economic and social outcomes* relevant to participating countries, including *individual outcomes* such as labour market participation and earnings, or participation in further learning and education, and *aggregate outcomes* such as economic growth, or increasing social equity in the labour market
- *Assess the performance of education and training systems*, and clarify which policy measures might lead to enhancing competencies through the formal educational system – or in the work-place, through incentives addressed at the general population, etc. and
- *Clarify relevant "policy levers"* (pp. 2–3, emphasis added).

The PIAAC objectives thus appear to comprise a 'human capital' approach, linked with social concerns (Evans et al., 2013).

In the framework used by OECD, Literacy, Numeracy and Problem-solving in technology-rich environments[2] are the three 'competencies' which PIAAC aims to measure. In the OECD's approach, *competencies* are:

> internal mental structures, i.e. abilities, capacities or dispositions embedded in the individual […] Although cognitive skills and the knowledge base are

> critical elements, it is important not to restrict attention to these components of a competence, but to include other aspects such as motivation and value orientation. (PIAAC Numeracy Expert Group, 2009, p. 10)[3]

Literacy is defined in PIAAC 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. (OECD, 2013b, p. 21)

Numeracy is defined for the purposes of designing the items for PIAAC 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. (OECD, 2013b, p. 26)

This is put forward as a basis for conceptualising mathematical thinking in context. However, in order to *operationalise* numeracy, the idea of *numerate behaviour* is developed, that is:

> the way a person's numeracy is manifested in the face of situations or contexts which have mathematical elements or carry information of a quantitative nature. […] inferences about a person's numeracy are possible through analysis of performance on assessment tasks designed to elicit numerate behaviour. (PIAAC Numeracy Expert Group, 2009, p. 10)

This led to specifying the following dimensions of "numerate behaviour" (or 'task characteristics') that can be used to guide the construction of assessment tasks:

- *context* (four types): personal, work-related, society and community, education and training
- *cognitive strategy* or *response* (three main types): identify/locate/access (information); act on/use; interpret/evaluate
- *mathematical content* (four main types): quantity and number, dimension and shape, pattern and relationships, data and chance, and
- *representations* (of mathematical/statistical information): e.g. text, tables, graphs.[4]

Each Numeracy item can be categorised on these four dimensions, along with its estimated difficulty ('ability level'); see (OECD, 2013a, pp. 26–28).

PIAAC also aims to produce affective and other contextual data that can be related to the respondent's performance. This includes demographic and attitudinal information in a Background Questionnaire (BQ), and self-report indicators on the respondent's use of, and need for, job-related skills at work.[5]

Each country has interviewed at least 5,000 adults, normally 16–65 years of age. PIAAC's default method of survey administration is by laptop computer,[6] although paper-based testing was used in IALS/ALL (and PISA up to now). As indicated above, this facilitates the use of *adaptive routing*.

Understanding PIAAC's Conceptual Framework and Methodology

In seeking to understand PIAAC and other adult skills surveys and their results, I consider how the interpretation of such studies needs to be related to their conceptual bases and methodological choices, as well as to arguments and decisions about presenting, reporting and reconceptualising them (e.g. Hamilton & Barton, 2000; Radical Statistics Education Group, 1982; Tsatsaroni & Evans, 2013).

Generally, surveys rely on aspects of the research design, responding to reasonably well-understood criteria of validity, to enhance and to monitor the measurement and sampling procedures. It is important for literacy and numeracy researchers, teachers and policy makers to be able to consider these when the results of a survey are presented and discussed. Here I consider the following likely effects of certain design features of the survey, and their realisation in the field:

- the *content validity* of the definitions of literacy, or of numeracy and numerate behaviour ('types' or categories of items, as above)
- the *measurement validity* of the items presented, including the administration and scoring procedures ('qualities' of items)
- the *reliability* of the measurement procedures, and
- the *external validity*, or representativeness, for the national population of interest, of the results produced from the sample (see Evans, 1983, for a fuller discussion).

In my discussion below, I will be referring to PIAAC Numeracy to explain these issues, but the same principles apply for Literacy.

Content Validity

Content validity refers to the extent to which a measure represents all aspects of a given concept. The definition of numeracy used by PIAAC (and, earlier, ALL) is based on the four dimensions of numerate behaviour stipulated above: *context, content, response, representation*. Each item can be categorised on these four dimensions, and the proportion of items falling into each category can be controlled over the whole set of items, so as to make the operational definition of numerate behaviour more explicit, and the content validity of the overall set of items more open to scrutiny. In PIAAC Numeracy, the proportion of items falling into each category of mathematical content, context, and response is controlled (OECD, 2013b, p. 28). This allows test designers to stipulate the proportions of the items that are from each type of each key dimension, and from different levels of difficulty[7] – for example, the proportion of 'data and chance' items of moderate difficulty.

Nevertheless, in an international survey, this provides a transnational definition, and one needs to question how well it 'fits' the lives of adults in any particular country. Indeed, the four types of *context* (Personal, Work-related, Society and community, Education and training) are *under-specified*: they are rather too general

to refer to any actual specific social practice or social context in which *any particular respondent* might engage, in everyday life.

Measurement Validity

What I call here 'measurement validity' refers to the extent to which the responses to the set of items administered to a respondent actually capture what the conceptualisation of numeracy specifies; this will depend on the *actual range of items used*. As with most large-scale educational assessments, the full set of the items used is not made public while the survey is on-going.[8]

Nevertheless, careful reading of the OECD publications allows us some insights into the Numeracy items used. All 56 actual numeracy items are categorised as to Content, Cognitive strategy and Context in the Numeracy Item Map (OECD 2013b, p. 66, Table 4.3). Five of these items, one from each of the levels from 'below level 1' to level 4, are described in more detail (OECD 2013a, pp. 77–78). And three numeracy 'sample items', not used in PIAAC but similar to items actually used, are published in OECD (2013b, pp. 28–30).

The latter sample of three 'PIAAC-like' items was published to represent the more than 50 that might potentially be presented to any PIAAC respondent. Like any sample, of course, these three items cannot represent the full range of combinations of Content, Context, Cognitive strategy, and Difficulty levels. Nevertheless, it may be useful to consider them briefly here, since they give some specificity to the more general characterisation of numeracy in the survey discussed above. For one of the items, the mathematical content is framed by Personal or Work-related contexts; for the other two, Society and community contexts[9]. They combine realistic images of the problem at hand and school-like test rubrics, providing the questions that need to be answered, presumably by applying the correct mathematical procedures; see OECD (2013b, pp. 28–30).

In any particular country, we can ask how well these sorts of tasks – such as making precise readings from the appropriate temperature scale (as in item 2), or detecting changes in a time series graph of live births (as in item 1) – might represent adults' social practices and everyday lives in that country. We should also ask whether tasks such as these would tap or encourage what we would consider as mathematical thinking about potentially challenging tasks. Sample item 3, which asks for a calculation of the number of wind turbines needed to replace the output of one decommissioned nuclear power station in Sweden certainly appears to represent a more challenging task for most adults in many of the countries surveyed by PIAAC in the current round.

Measurement validity also requires procedures designed for the administration of the survey to be standardised in advance across all countries, e.g. design specifications of the laptops and software to be used, and rules for access to calculators and other aids[10]. As with any survey, full appreciation of the validity of procedures requires

assurance of how these procedures are followed in the field. This is even more crucial when results are compared across countries using different fieldwork teams.

External Validity

External validity includes the question of the representativeness of the sample for the population of interest; thus, the 5,000 or more adults (usually aged 16–65) selected for the sample in each country need to represent the population of that country. We can scrutinise, for any participating country, the sample design and other key aspects, such as the incentives offered to those selected for the sample to encourage their participation in the survey. Again, judgments about the effectiveness of these procedures depend partly on knowledge of the actual field practices.

However, it is important to realise that any result from such a sample, whether the mean score for a country, or a difference (e.g. by gender) in the percentages of items correct, is only an *estimate for the corresponding population value* **(**of the mean or the size of the difference in percentages). The population value for the whole country is what we would really like to know about – but this is not possible with certainty, since we only 'know' about the (hopefully 'representative') sample that our methods have chosen. Other samples, chosen in an equally 'correct' way, would (almost certainly) give different results. So virtually every numerical result that we produce with a sample survey cannot be considered *exact*, but should have a 'tolerance', *a margin of error*, on either side of the sample-based estimate. In this way we can be reasonably 'confident' that the population value (though its exact value is unknown) will be within a specified interval.[11]

Thus, if we consider Australia's average score in Numeracy, it is estimated as 268 points (267.6 to one decimal place), based on results (OECD, 2013a, p. 263, Table 2.6a) from a sample of 7,428 adults (OECD 2013b, p. 54). But this estimate of the average score of the entire population of adults 15–74 (about 16 million – see next section) cannot be exact (see above). Thus, a 95 percent confidence interval for the population average for Numeracy in Australia will be between 266 and 270 points.[12]

Sometimes, this use of confidence intervals leads to 'surprises'! For example, in the ranking by average Numeracy score, the first four countries are (to one decimal place):

Japan 288.2 Finland 282.2 Belgium 280.4 Netherlands 280.3

This appears to be a very neat ranking, except that Belgium and Netherlands are just about equal. However, if we produce 95 percent confidence intervals for each country's score, in order to be reasonably confident that we have allowed for sampling variation, we get the following intervals for each score (rounded to nearest scale-point):

Japan 287 to 290 Finland 281 to 284 Belgium and Netherlands 279 to 282

Thus we can see that while Japan is still clearly 'ahead', the estimates for Finland, Belgium and Netherlands cannot be clearly separated, since their confidence intervals overlap: our neat ranking of countries looks much less clear-cut when we allow for sampling variation! Therefore, the OECD publishes analyses that allow for sampling variation.[13]

Reliability

The comparability of test administration across countries and across interviewers, and especially assuring the use of the same standards and practices in marking, has been a problem with past international surveys. Computer presentation and marking of test items will help greatly with *reliability*, the assurance that the survey will produce the same or very close results, if it were to be repeated, using the same procedures. But it may tend to undermine *content validity*, if it reduces the range of types of question that can be asked; for example, it is difficult to produce an item that asks a respondent to *give reasons for his/her answer*, if the item is to be presented and marked by computer. This trade-off between content/measurement validity and reliability is a well-known dilemma in research design.

Further, the strengthening of reliability may lead to concerns about loss of another aspect of *external validity*, namely *ecological validity*, i.e. whether the setting of the research is representative of those to which one wishes to generalise the results. For example, the on-screen presentation of tasks may not be representative of the settings in which respondents normally carry out tasks involving numeracy, and so may not facilitate their 'typical' thinking and behaviour responses. Again, similar dilemmas arise for much educational assessment – but must be considered afresh in understanding PIAAC results.

Beyond Methodology

This discussion of issues related to various aspects of the validity of the survey shows the importance of sound research design – and also of the way field work is accomplished. However, a number of key issues in interpreting the uses and effects of the survey go beyond the technical issues around methodological validity (e.g. Radical Statistics Education Group, 1982). They include the way that the survey's measured scores are *interpreted*/*reconceptualised* in presentations and reports of various interested parties. This aspect is of course not under the complete control of the survey's sponsors: for example, the media and certain national interests have often offered conflicting interpretations ('spin') of results of international surveys. Understanding these processes requires an appreciation of the policy context and the ideological debates that surround the reception of results in a particular country, as well as of the global education policy discourse.

Several examples can be given of the need for care and scepticism about the reporting and interpretation of these results; see e.g. *European Educational Research*

Journal (2012), on the way that PISA results are reported and used, and in particular, Carvalho on the "plasticity of knowledge" (2012, pp. 180–183). One problem is that an adult's performance on one of the subtests such as Numeracy cannot simply be expressed as the 'proportion correct' – since adaptive routing means that some respondents were presented with 'harder' items, and some 'easier'. So Item Response Modelling is used to ('psychometrically') estimate a standardised score – for PIAAC, scores are estimated in the range zero to 500, with standard deviation 50. Then, the numerical score is commonly related to one of five general 'levels' of Literacy or Numeracy to make it meaningful.

Now, this may well be more informative than simply reporting the percentage of adults in a country that are categorised as 'literate' or not, as was the case before OECD (and other) international or national surveys. But as in all such surveys, there is debate about use of a simple and *one-dimensional* characterisation of an adult's Numeracy or Literacy. For example, Gillespie (2004) referring to the first UK Skills for Life survey (done using a similar methodology to PIAAC) notes: "The findings confirm that for many, being 'at a given level' is not meaningful for the individual, as levels embody predetermined assumptions about progression and relative difficulty" (p. 1). Part of this scepticism flows from the finding that many adults have different 'spiky profiles', due to distinctive life experiences (Gillespie, 2004, pp. 4–6). Thus, some adults may find items of type A Content (say, 'data and chance') more difficult than type B items (e.g. 'dimension and shape') – and others find the opposite.

Similarly, some policy-makers may attempt to stipulate 'the minimum level of numeracy (or literacy) needed to cope with the demands of adult life' in their particular country – but this notion too is questionable; see Black and Yasukawa's (2014) discussion of current debates in Australia. Such generalising claims group together adults with different work, family and social situations, and different literacy or numeracy 'demands' on them.

These sorts of concerns about validity and interpretation are shared by users of all surveys including assessments, especially those that aim to make comparisons across countries, or over time. Nevertheless, such questions must be assessed for any survey, where results aim to inform policy or practice.

SOME FURTHER RESULTS FOR PIAAC FROM AUSTRALIA

A preliminary summary of the methodology and results from Australia was made available in February 2013, by the contractor, the Australian Bureau of Statistics (ABS, 2013a). This provided an indication of the sorts of results that became available in each of the participating countries from October 2013. Here I give three examples.

Figure 1 shows the proportions of Australian adults at different skills levels. Approximately 7.3 million (44 percent) Australians aged 15 to 74 years had Literacy skills at Levels 1[14] and 2, a further 6.4 million (39 percent) at Level 3 and 2.7 million (17 percent) at Levels 4/5. For the Numeracy scale, approximately 8.9 million

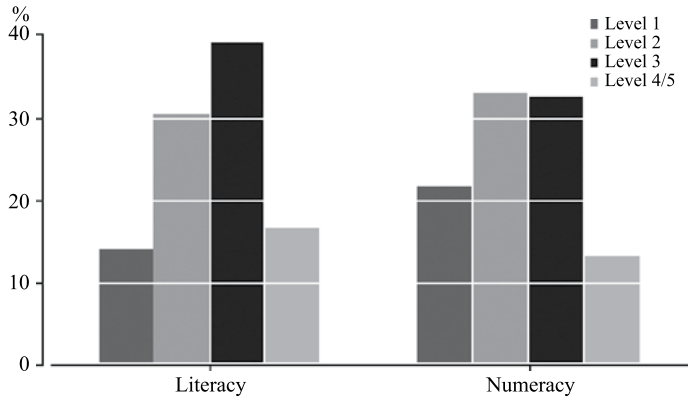

Figure 1. Overall results from PIAAC for Literacy and Numeracy: Australia, 2013. Source: Australian Bureau of Statistics (2013a)

(55 percent) Australians were assessed at Levels 1 and 2, 5.3 million (32 percent) at Level 3 and 2.1 million (13 percent) at Level 4/5. One could also compare Literacy and Numeracy levels for subgroups, e.g. residents of different Australian states. Thus, for Numeracy, Australian Capital Territory recorded the highest proportion of adults at Level 4/5 (23 percent). One can also ask about gender differences, frequently of interest in research like this; see Figures 2a and 2b.

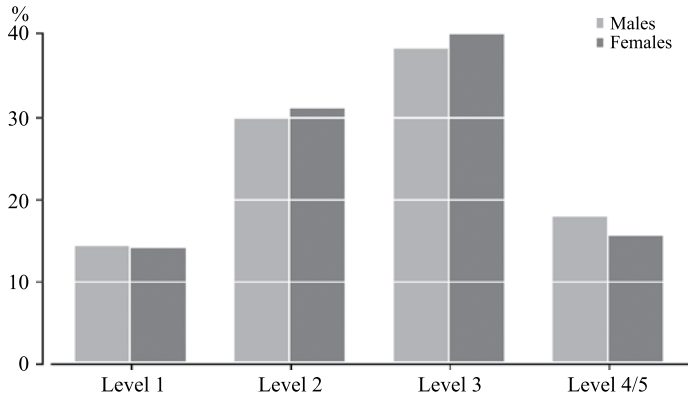

Figure 2a. Proportion at each PIAAC Literacy level, by sex: Australia 2013. Source: ABS (2013a)

In Figure 2a, there appears to be little difference in the proportion of males and females at each level of the Literacy scale. However, when we consider Numeracy results in Figure 2b we see that a higher proportion of males (17 percent) attained

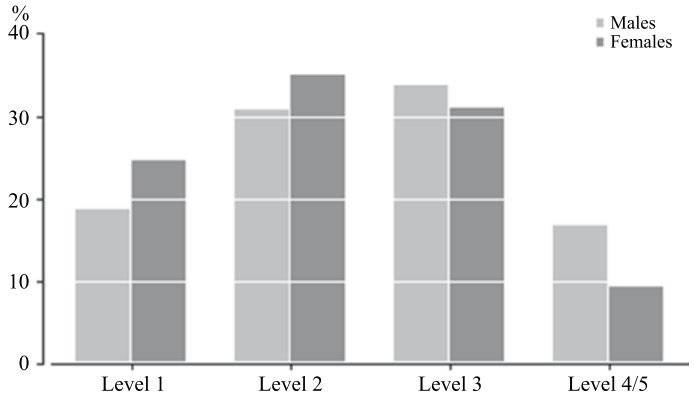

Figure 2b. Proportion at each PIAAC Numeracy level, by sex: Australia 2013. Source: ABS (2013a)

scores at Levels 4/5 compared with females (9 percent); the difference in those attaining Level 3 or above, about 50 percent of males compared with 42 percent of women, appears less striking, but still noteworthy.

One of the concerns of policy-makers is that younger generations should exhibit a higher level of skills than older people; otherwise there are anxieties about a national 'decline in skills', and loss of competitiveness over time. Hence, there has been much interest in the PIAAC countries in performance variations across the age range. Such differences afford some insight into the policy problem – though it is limited (see end of this section). We can consider Figures 3a and 3b, which show results from Australia where the age group surveyed was 15–74 (wider than the 16–65 range studied in most other countries).

In Figures 3a and 3b, we can see that Literacy and Numeracy scores show an increase in assessed scores from the youngest age group, reaching a peak in the middle years (late 20s to early 40s), and then declining from the late 40s. For example, the percentage of people (males and females) with Literacy skills at Level 3 or above was 54 percent for people aged 15 to 19 years, 63 percent for people aged 25 to 34 years, 54 percent for people aged 45 to 54 years and 28 percent for people aged 65 to 74 years (ABS, 2013b). The percentage of people with Numeracy skills at Level 3 or above was 42 percent for people aged 15 to 19 years, 51 percent for people aged 25 to 34 years, 45 percent for people aged 45 to 54 years and 24 percent for people aged 65 to 74 years (*ibid*.).

In Literacy younger women outscored younger men, though "there was no [statistically] significant difference", while "(f)ewer older women had literacy skills at Level 3 or above, than their male counterparts" (ABS, 2013b). For Numeracy "more men were assessed at Level 3 or above than women at all ages, but the difference, which was ten percentage points or higher for older ages, was lower for

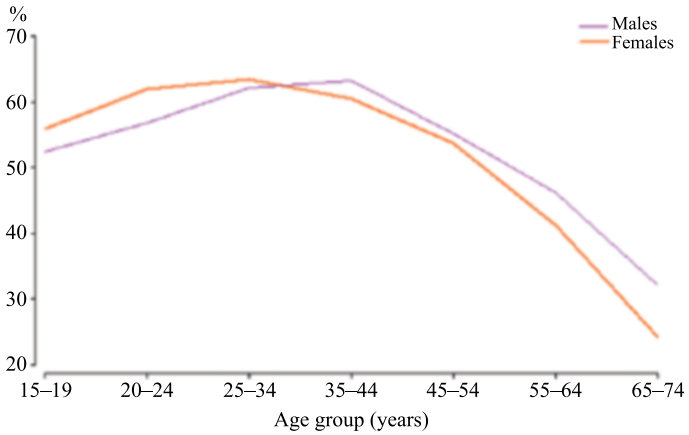

Figure 3a. Proportion at Literacy level 3 or above, by sex and age group, Australia, 2013. Source: ABS (2013b)

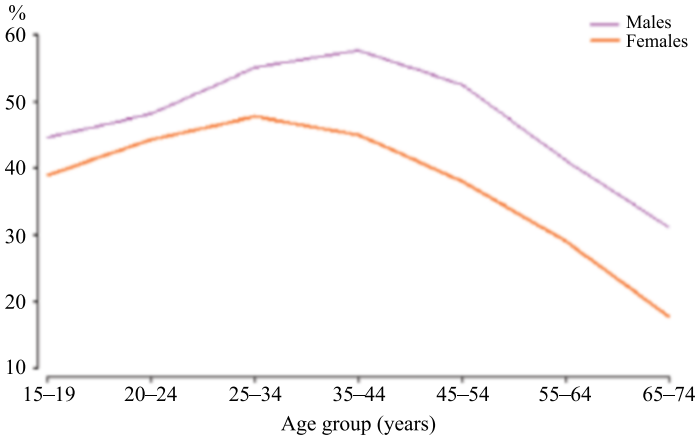

Figure 3b. Proportion at Numeracy level 3 or above, by sex and age group, Australia, 2013. Source: ABS (2013b)

younger ages." Overall, one could say that younger women had *relatively higher* scores in both Literacy and Numeracy (compared to men) than older women did (ABS, 2013b).

These debates about differences in performance by age (and gender) which were evident around previous international (and national) performance studies show no signs of diminishing (e.g. Reder, 2009). The evidence from PIAAC is limited as it is, so far, a *cross-sectional* (one-off comparative), and not a *longitudinal*,

survey.[15] The latter design responds to the need, in these discussions, to separate *age factors*, *cohort factors* and *historical-contextual* ones. For example, an individual's skills may increase or decrease as they age – or they may not. And later cohorts in most countries normally have had on average more years of formal education. And, at the same time, some groups of adults in particular enterprises in particular countries may have more or fewer opportunities to develop their skills at work.

DISCUSSION: POSSIBLE EFFECTS OF INTERNATIONAL SURVEYS AND 'COUNTERVAILING FORCES'

In considering the possible effects of international surveys on the teaching and learning of adult literacy and numeracy, we can draw on Basil Bernstein's analysis (2000) of the structuring of pedagogic institutions and discourses. This analysis and his focus on changing forms of educational knowledge and practices can illuminate shifts in the mode of governance of educational policy, in which international surveys play a role (Tsatsaroni & Evans, 2013). This framework can also be used to critique a globally promoted type of pedagogic discourse, which asserts adults' need for certain rather generic skills, and countries' need to assess these in a comparative way.

The international adult studies, like IALS, ALL and PIAAC, have no systematically thought out curriculum associated with them (unlike TIMSS and PISA). Yet the existence of such a 'curriculum' is arguably implied in the definitions of literacy and numeracy[16], the descriptions of 'levels' of performance, and the use (for numeracy) of existing classifications of mathematical content. Tsatsaroni and Evans (2013) earlier thought there was "a strong possibility that PIAAC could reinforce this type of pedagogic discourse, and the surveys could tend to work as an exemplary curriculum type which indirectly would prescribe what knowledge the adult populations *in all societies* should value, strive to acquire, and demonstrate" (Tsatsaroni & Evans, 2013, p. 178, emphasis added). Indeed, Christine Pinsent-Johnson's more recent paper (2015) shows that this "possibility" has already materialised. She demonstrates how texts and textual devices, including international assessment test tasks and descriptions of performance 'levels', developed in the context of an international skills assessment initiative like PIAAC (or IALS), are "transposed" into the context of adult literacy education, as part of the Essential Skills in Canada, a "competency-based occupational standards framework". She describes how "once the texts enter into local programmes via the Essential Skills, […] they mediate how literacy is conceptualised, taught and valued" (Pinsent-Johnson, 2015, pp. 201–202). She concludes that her textual analysis reveals, *inter alia*, the following consequences for teaching and learning: "the paucity of mechanisms in the test task methodology that can be used to inform educators about actual literacy uses in people's daily lives, and its developmental trajectory", and "the displacement and disestablishment of literacy learning expertise" (p. 202).

There are a number of other possible effects of such performance surveys, which may come to represent 'high stakes' for adults and the countries involved. An obvious negative effect is the pathologising of countries which do not 'perform' to standards – not necessarily by the survey's sponsors, but by sections of the media, political parties, and new educational agencies, such as national assessment bodies. (cf. "PISA shock", discussed in *European Educational Research Journal*, 2012).

The emerging discourse supported by international surveys may also have effects on teachers', learners', researchers' and citizens' ways of understanding adult literacy and numeracy.[17] Knowledge comes to be seen as generic skills, flowing from a *decontextualised imagining* of the adult's everyday practices (Hamilton, 2012). This may result in differential access, across social groupings and of countries, to the principles of thinking that disciplinary or professional forms of knowledge can provide (Tsatsaroni & Evans, 2013).

Now, disciplinary knowledge, say in mathematics, can also be understood as '*powerful knowledge*' (Young, 2010) – or as 'big ideas' in mathematics education (Lerman, Murphy, & Winbourne, 2013) – that is, as ideas that have rich applicability in a range of fields. One example is the idea of *conditional probability*. This idea occurs under many guises: as 'having the *right denominator* for your proportions'; or in reporting research results (e.g. percentage of items correct) *for the appropriate population*; or in appreciating the difference between *the probability of testing positive for x, given that you have disease x* – and *the probability of having disease x, given that you test positive for x*, which is vital in understanding medical test results (Gigerenzer, 2003; O'Hagan, 2012.)

At the same time, it is worth investigating whether international surveys might afford opportunities for further research. Though results are anonymous at individual level, there is potential for relating performances of *categories of respondents* – to demographic and attitudinal data from the Background Questionnaire, and/or further information available on numeracy related practices and 'use of skills' at work. These studies may provide additional ways to study established topics, such as affect (attitudes) among adult learners (Evans, 2000). They may also provide a context for certain types of national studies, or local qualitative studies, to supplement or to probe Background Questionnaire results; for example to investigate why residents of the Australian Capital Territory might have recorded the highest proportion of adults at Level 4/5 for numeracy (23 percent; see above). There are also some examples of use of results from earlier international surveys, e.g. PISA and TIMSS, to study wider educational and social questions (see e.g. Kanes, Morgan, & Tsatsaroni, 2014; Meyer & Benavot, 2013).

In addition, OECD policy is to make available, on their website, datasets from PIAAC – and software for data analysis – for research purposes. This was done at the same time as the release of the results in October 2013. Thus, resources for researching interesting questions suggested by the preliminary results are now more accessible than before.

We can also look to alternative research programmes to produce critical resources to help with asserting the value of alternative conceptions of educational knowledge, and with appreciating developments in adult educational policy issues, including literacy and numeracy. From within adult numeracy, we can illustrate ways to challenge the currently dominant ideas of numeracy and adult skills. For example, Coben and her colleagues have challenged the conventional 'deficit' characterisation of practising nurses' numeracy, and argued that often the high-stakes testing programmes used for this deployed instruments which lacked reliability, validity, and authenticity (e.g. Coben, 2010). Hoyles, Noss, Kent and Bakker (2010) go beyond a narrow definition of numeracy to develop a richer conception of 'Techno-mathematical Literacies' (TmLs), informed by the affordances, flexibilities and demands of information technologies, and document its use by middle ranking UK professionals, in decision-making in specific workplaces. Mullen and Evans (2010) describe demands on citizens' numerate thinking and learning, emphasising the social supports made available by government and other institutions, in coping with the 2009 euro conversion in the Slovak Republic. Gelsa Knijnik and her colleagues describe work with the Landless Movement in Brazil, facilitating their learning to recognise, to compare, and to choose appropriately from academic and/or 'local' knowledges, in carrying out their everyday practices (e.g. Knijnik, 2007).

Powerful knowledges of all these kinds can empower on a broader social basis, through knowledge located in the disciplines, professional practice, or other established practices of adults' 'lived experience'. The aim of educational researchers must be to support the development of potentially powerful knowledge (Young, 2010), like numeracy and literacy, and to prevent their being reduced to narrow competencies.

ACKNOWLEDGEMENTS

I thank Anna Tsatsaroni, Tine Wedege and Keiko Yasukawa for useful discussions supporting the arguments in this chapter. I also thank members of the PIAAC Numeracy Expert Group, and colleagues in Adults Learning Mathematics – a Research Forum, for stimulating exchanges on adults' mathematics education and numeracy, over many years. Appreciation is due also to colleagues in the Radical Statistics Group for valuable discussions on methodology in education and the social sciences, over the last 40 years.

NOTES

[1] The Programme for International Student Assessment (PISA) is sponsored by OECD, while the Trends in International Mathematics and Science Study (TIMSS) is sponsored by the International Association for the Evaluation of Educational Achievement (IEA).

[2] In this chapter, I use upper case for proficiencies, as measured in the PIAAC survey (e.g. 'Literacy'); and lower case for the concept, as used by researchers or the general public (e.g. 'literacy').

[3] The reader should refer to this 2009 document for more detailed discussion on the numeracy assessment framework, and to OECD (2013a), concerning all three competencies in PIAAC.
[4] Literacy items are characterised by a similar, but not identical, set of dimensions (OECD, 2013a, pp. 21–22).
[5] See OECD (2013b, pp. 27–46 and 114, Table B1) for the BQ's conceptual framework and Central Statistical Office, Ireland (2013) for a copy of the BQ.
[6] Respondents are presented with initial computer-based tasks; anyone uncomfortable with these takes an alternative pencil-and-paper version.
[7] These levels of difficulty are estimated by the Item Response Modelling procedures; see below.
[8] Round 2, including a further nine countries (e.g. New Zealand, Singapore and Indonesia), is completing fieldwork in 2014–15, and reporting in 2016.
[9] The OECD Framework document indicates that the overall distribution of Numeracy items included by contexts was: Personal – 45 percent; Work-related – 23 percent; Society and community – 25 percent: Education and training – seven percent (OECD, 2013b, p. 28).
[10] Respondents in the first round of PIAAC, completed in 2011–12, were supplied with hand held calculators and rulers with metric and imperial scales, for use during the interview.
[11] In this chapter I use an intuitive notion of 'confidence', ranging between zero percent and 100 percent. The margin of error depends on the degree of 'confidence' desired in the estimate, but is normally two standard errors for a 95 percent confidence interval.
[12] For the means and standard errors (SEs) used for calculations here, see Table 2.6a in OECD (2013a, p. 263).
[13] For example, Figure 2.6a in OECD (2013a, p. 80) shows that, even if country A appears two or three 'positions above' country B in the rankings, their results may nonetheless be effectively indistinguishable ('not statistically significantly different'), once we allow for sampling variation. We can see this in the fact that the 'superiority' of, say, Finland, over the Netherlands in terms of average Numeracy score is only apparent.
[14] Later analyses distinguished those at Level 1 from those 'below Level 1' (e.g. OECD, 2013c).
[15] However, as indicated in the Introduction, there are aspirations to repeat PIAAC in at least some countries over time, and some longitudinal insights can be gained by linking PIAAC results to those from IALS and ALL in certain countries.
[16] The definition of numeracy outlined earlier pointed to the abilities and competencies required "in order to engage in and manage the mathematical demands of a range of situations in adult life" (PIAAC Numeracy Expert Group, 2009, pp. 20ff).
[17] And lifelong learning more generally (Evans, Wedege, & Yasukawa, 2013).

REFERENCES

Australian Bureau of Statistics (ABS). (2013a). *Programme for the International Assessment of Adult Competencies, Australia, 2011–12: Preliminary findings, 15/2/2013*. Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4228.0~20112012~Main%20Features~Preliminary%20findings~99

Australian Bureau of Statistics (ABS). (2013b). *Programme for the International Assessment of Adult Competencies, Australia, 2011–12: Latest issue 9/10/2013*. Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4228.0Main+Features202011-12

Bernstein, B. (2000). *Pedagogy, symbolic control and identity: Theory, research, critique* (Revised version). Oxford: Rowan & Littlefield Publishers.

Black, S., & Yasukawa, K. (2014). Level 3: Another single measure of adult literacy and numeracy. *Australian Educational Researcher*, *41*(2), 125–138.

Brown, P., Halsey, A. H., Lauder, H., & Wells, A. S. (1997). The transformation of education and society: An introduction. In A. H. Halsey, H. Lauder, P. Brown, & A. S. Wells (Eds.), *Education: Culture, economy, and society* (pp. 1–44). Oxford: Oxford University Press.

Carvalho, L. M. (2012). The fabrications and travels of a knowledge-policy instrument. *European Educational Research Journal*, *11*(2), 172–188. Retrieved from http://dx.doi.org/10.2304/eerj.2012.11.2.172

Central Statistical Office, Ireland. (2013). *Surveys and methodology: PIAAC*. Retrieved from http://www.cso.ie/en/media/csoie/surveysandmethodologies/surveys/education/piaac/documents/PIAACSampleTasks.pdf

Coben, D. (2010). At the sharp end of education for an ethical, equitable and numerate society: Working in a safety-critical context – Numeracy for nursing. In U. Gellert, E. Jablonka, & C. Morgan (Eds.), *Proceedings of the sixth international mathematics education and society conference* (pp. 9–22). Berlin, Germany: Freie Universität.

European Educational Research Journal. (2012). Governing knowledge? PISA in focus. *European Educational Research Journal*, *11*(2). (Special Issue; L. M. Carvalho, S. Grek, E. Mangez, & X. Pons (Eds.)). Retrieved from http://www.wwwords.co.uk/eerj/content/pdfs/11/issue11_2.asp

Evans, J. (1983). Criteria of validity in social research: Exploring the relationship between ethnographic and quantitative approaches. In M. Hammersley (Ed.), *The ethnography of schooling* (pp. 173–201). Driffield: Nafferton Books.

Evans, J. (2000). *Adults' mathematical thinking and emotions: A study of numerate practices*. London: Routledge Falmer.

Evans, J., Wedege, T., & Yasukawa, K. (2013). Critical perspectives on adults' mathematics education. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 203–242). New York, NY: Springer.

Gigerenzer, G. (2003). *Reckoning with risk – Learning to live with uncertainty*. London: Penguin.

Gillespie, J. (2004). *The "Skills for Life" national survey of adult numeracy in England. What does it tell us? What further questions does it prompt?* Paper for Study Group 6: Adult and Lifelong Education, 10th International Congress on Mathematical Education (ICME-10).

Grek, S. (2010). International Organisations and the shared construction of policy 'problems': Problematisation and change in education governance in Europe. *European Educational Research Journal, 9*(3), 396–406.

Guadalupe, C. (2015). How feasible is it to develop a culturally-sensitive large-scale standardised assessment of literacy skills? In M. Hamilton, B. Maddox, & C. Addey (Eds.), *Literacy as numbers: Researching the politics and practices of international literacy assessment* (pp. 111–128). Cambridge: Cambridge University Press.

Hamilton, M. (2012). *Literacy and the politics of representation*. London: Routledge.

Hamilton, M., & Barton, D. (2000). The international adult literacy survey: What does it really measure? *International Review of Education*, *46*(5), 377–389.

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

Kanes, C., Morgan, C., & Tsatsaroni, A. (2014). The PISA mathematics regime (of rationality) and the ethics of self. *Educational Studies in Mathematics. Special Issue on Social Theory in Mathematics Education*, *87*(2), 145–165.

Knijnik, G. (2007). Brazilian peasant mathematics, school mathematics and adult education. *Adults Learning Mathematics – An International Journal*, *2*(2), 53–62.

Lerman, S., Murphy, B., & Winbourne, P. (2013). *Big ideas in teacher knowledge and mathematical pedagogy*. London: London South Bank University, unpublished.

Meyer, H.-D., & Benavot, A. (Eds.). (2013). *PISA, power and policy: The emergence of global educational governance*. Oxford: Symposium Books.

Moore, R., & Jones, L. (2007). Appropriating competence: The competency movement, the new right and the 'culture change' project. In R. Moore (Ed.), *Sociology of knowledge and education* (pp. 130–144). London: Continuum.

Mullen, J., & Evans, J. (2010). Adult numeracy in the context of economic change: The conversion to the Euro in the Slovak Republic. In G. Griffiths & D. Kaye (Eds.), *Numeracy works for life*: *Proceedings of the 16th International Conference of Adults Learning Mathematics – A Research Forum (ALM) incorporating the LLU+ 7th National Numeracy Conference* (pp. 189–200). London: ALM/ London South Bank University.

O'Hagan, J. (2012). (When) can we trust ourselves to think straight? – and (when) does it really matter? In T. Maguire, J. J. Keogh, & J. O'Donoghue (Eds.), *Mathematical eyes: A bridge between adults, the world and mathematics: Proceedings of the 18th International Conference of Adults Learning Mathematics – A Research Forum (ALM)* (pp. 305–315). Dublin: ALM & ITT Dublin.

Organisation for Economic Co-operation and Development (OECD). (2007). *Qualifications systems: Bridges to lifelong learning*. Paris: OECD Publishing. Retrieved from http://www.oecd.org/edu/skills-beyond-school/38465471.pdf

Organisation for Economic Co-operation and Development (OECD). (2013a). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD. Retrieved from www.oecd.org/site/piaac/#d.en.221854

Organisation for Economic Co-operation and Development (OECD). (2013b). *The survey of adult skills: Reader's companion*. Paris: OECD. Retrieved from www.oecd.org/site/piaac/publications.htm

Organisation for Economic Co-operation and Development (OECD). (2013c). *Country note – Australia: Survey of adult skills – First results*. Paris: OECD. Retrieved from www.oecd.org/site/piaac/Country%20note%20-%20Australia_final.pdf

Ozga, J. (2009). Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy*, *24*(2), 149–162.

PIAAC Numeracy Expert Group. (2009). *PIAAC numeracy: A conceptual framework* (OECD Education Working Paper No. 35). Paris: OECD Publishing. Retrieved from http://www.oecd.org/officialdocuments/Displaydocumentpdf?cote=edu/wkp(2009)14&doclanguage=en

Pinsent-Johnson, C. (2015). From an international adult literacy assessment to the classroom: How test development methods are transposed into curriculum. In M. Hamilton, B. Maddox, & C. Addey (Eds.), *Literacy as numbers: Researching the politics and practices of international literacy assessment* (pp. 187–205). Cambridge: Cambridge University Press.

Radical Statistics Education Group. (1982). *Reading between the numbers: A critical guide to educational research*. London: BSSRS Publications. Retrieved from http://www.radstats.org.uk/books/ReadingBetweenNumbers.pdf

Reder, S. (2009). The development of literacy and numeracy in adult life. In S. Reder & J. Bynner (Eds.), *Tracking adult literacy and numeracy skills: Findings from longitudinal research*. London: Routledge.

Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. London: Routledge.

Schleicher, A. (2008). PIAAC: A new strategy for assessing adult competencies. *International Review of Education*, *54*(5–6), 627–650. Retrieved from http://www.oecd.org/dataoecd/48/5/41529787.pdf

Tsatsaroni, A., & Evans, J. (2013). Adult numeracy and the totally pedagogised society: PIAAC and other international surveys in the context of global educational policy. *Educational Studies in Mathematics, Special Issue on Social Theory in Mathematics Education, 87*(2), 167–186.

Walker, J. (2009). The inclusion and construction of the worthy citizen through lifelong learning: A focus on the OECD. *Journal of Education Policy*, *24*(3), 335–351.

Young, M. F. D. (2010). Alternative educational futures for a knowledge society. *European Educational Research Journal*, *9*(1), 1–12.

Jeff Evans
Middlesex University