

New PIAAC Results: Care Is Needed in Reading Reports of International Surveys

Jeff Evans

Middlesex University, London NW4 4BT, UK
<j.evans@mdx.ac.uk>

Abstract

Results from the Survey of Adult Skills, also known as PIAAC (Programme for the International Assessment of Adult Competencies), were recently made available for 24 participating countries. PIAAC involves several developments in relation to the earlier international “adult skills” surveys (IALS in the 1990s and ALL in the 2000s), notably the use of computer administration of the survey. In this paper, I focus on understanding these studies, by considering conceptual issues, methodological validity of research design and execution, and presentation of results. I consider several of the sample items for numeracy published by OECD (2012). And I discuss illustrative results from Australia made available in February 2013, by the Australian Bureau of Statistics. The paper shows when and how to be sceptical when reading international survey reports. It also opens up questions concerning the relevance of the results, and the other types of research that may be needed, in different national and local contexts.

Key words: adult skills, international assessment, and mathematics

Introduction

From October 2013, results from PIAAC (Programme for the International Assessment of Adult Competencies) for 24 participating countries have been available (OECD, 2013a, 2013b). PIAAC aims to provide information as an international comparative survey, successor to IALS (during the 1990s) and ALL (2000s), and it has many similarities with national studies, such as Skills for Life in the UK. Unlike the school level surveys (TIMSS, PISA), which gain access to “captive populations” in schools, PIAAC needs to use a combination of household survey and educational testing methodologies. It involves developments from the adult earlier studies, in several ways.

The first round covers a greater range of countries (24, two thirds of which are EU members, with the rest from North America, East Asia and Australia) – though all are advanced industrial economies. It focuses on three domains or “competencies” – Literacy, Numeracy, and now Problem Solving in Technology Rich Environments (PSTRE). It uses computer administration, which has a number of consequences, in particular allowing *adaptive routing* of respondents (see Section 3), and making the survey results available more quickly and more accessibly. In addition, PIAAC has implemented a number of methodological and fieldwork improvements, for example, tighter specification and regulation of sampling and fieldwork standards than in previous international surveys (OECD, 2013b, pp. 47-61).

PIAAC is designed to be repeated, in order to build up time series data for participating countries. This longitudinal aspect would aim to facilitate the study over time of the correlations of the performance outcomes with relevant social or attitudinal variables.

In Section 2 I sketch international the policy context, including the conception of Lifelong Learning (LLL) promoted by the survey's sponsor, the OECD (Organisation for Economic Cooperation and Development). In Section 3, I describe the survey aims, and the underlying conception of numeracy. In Section 4, I consider how this conception is deployed in the measurement process, and other aspects of *methodological validity* that need to be considered for international performance surveys. I also focus on the need to consider the way that the survey results are *reported*, since this crucially affects the way “the findings” are perceived by various categories of readers. In Section 5, I discuss some illustrative results from Australia, and in Section 6, I return to focus on the effects of international surveys like PIAAC on the developing educational policy context worldwide.

The international policy context

Educational policy is currently being developed on a world-wide scale, with supranational organisations acting as key agencies for change. Increasing globalisation and competitive economic environments are leading national governments to seek competitive advantage – which is “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) seek to provide measures of a country's progress according to international standards.

The idea of Lifelong Learning (LLL) is central to the conceptualisation of adult numeracy (and literacy). In international policy debates, LLL has been much contested, e.g. between “humanistic” (learning for the whole person) and “economistic” (human capital) approaches (Evans, Wedege, & Yasukawa, 2013). In this connection, it is important to consider work done both within the UNESCO programmes (e.g. Guadalupe, 2013), and by the OECD.

Here I focus on the OECD, PIAAC's sponsor. OECD's view of LLL aims to promote the development of knowledge and competencies enabling each citizen to actively participate in various spheres of globalised social and economic life. It also promotes a broad view of the context of learning, and a weakening of the distinction between formal and informal education. At the same time, it emphasises 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). Some of the consequences of these commitments will be discussed below; see also Tsatsaroni & Evans (2013).

The European Union (EU) is working closely with OECD on PIAAC. For supra-national institutions like the EU, the area of Lifelong Learning 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, and 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. These practices allow countries to measure the *relative* success of their education systems and to shift policy orientations accordingly, while increasingly facilitating the role of these supra-national organisations themselves to be “governing by data” (Ozga,

2009). Overall, 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).

The PIAAC Survey

PIAAC’s main 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. (pp. 2-3).

The PIAAC objectives thus appear to comprise a “human capital” approach, coupled with social concerns (Evans, Wedege & Yasukawa, 2013).

In the framework used by OECD, *literacy*, *numeracy* and *problem-solving* in technology rich environments 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)

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. (PIAAC Numeracy Expert Group, 2009, pp. 20ff.)

This is put forward as a basis for conceptualising mathematical thinking in context. However, in order to *produce measures* of numeracy, the idea of *numerate behaviour* is put forward, 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” that can be used to guide the construction of assessment tasks:

- *context* (four types): everyday (or personal), work, society and community, further learning
- *response* (to mathematical task - 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

- *representations* (of mathematical / statistical information): e.g. in text, tables, and/or graphs¹.

Each item can be categorised on these four dimensions, along with its estimated difficulty.

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; see OECD (2013b) for the BQ's conceptual framework, and CSO, Ireland (2013) for a copy of the BQ.

Each country interviewed at least 5000 adults, normally 16-65 years of age. PIAAC's default method of survey administration is by laptop computer², although paper-based testing was used in IALS / ALL (and PISA up to now). This facilitates the use of *adaptive routing*, which estimates the "skill level" of the respondent from a few initial responses, and then administers more appropriate items (in terms of difficulty) throughout the rest of the interview.

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 decisions, as well as choices about presentation and reporting and arguments about the range of applicability of the findings (Tsatsaroni & Evans, 2013; Hamilton & Barton, 2000; Radical Statistics Education Group, 1982 / 2012).

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 likely effects of certain design features of the survey, and their realisation in the field, in terms of the following aspects of validity:

1. the *content validity* of the definitions of numeracy and numerate behaviour ("types" or categories of items, as described above)
2. the *measurement validity* of the items presented, including the administration and scoring procedures ("qualities" of items)
3. the *reliability* of the measurement procedures
4. 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.)

Content validity

I am using the term *content validity* in this paper to refer to the extent to which a measure represents all aspects of a given concept, as it is defined. The definition of numeracy used by PIAAC (and, earlier, ALL) is based on the four dimensions of numerate behaviour stipulated above. Each item can be categorised on these four dimensions, and the proportion of items falling into each category can be controlled over the whole scale, 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

¹ Literacy and PS-TRE items are characterised by a similar, though not identical, set of dimensions (OECD, 2013b, pp. 21-34).

² Respondents are presented with initial computer-based tasks; anyone uncomfortable with these takes an alternative pencil-and-paper version of the main tests.

each key dimension, and from different levels of difficulty³ – for example, the proportion of “data and chance” items of medium difficulty.

However, in an international survey, this can provide only a general, transnational definition, and one needs to question how well it “fits” adults’ lives in any particular country. For example, the four types of *context* (everyday / personal, work, society and community, further learning) can be specified only in a rather general way – they may or may not represent the repertoire of actual specific social practices or social contexts in which *any particular respondent* might engage, in his/her life. Thus we need to examine a set of items that a particular sample member might be asked to respond to.

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-going⁴. Nevertheless, four illustrative items are presented on several websites (e.g. CSO, Ireland, 2013), and in the Appendix.

This sample of four PIAAC or “PIAAC-like” numeracy items were published to represent the more than 50 that might potentially be presented to any PIAAC respondent (OECD, 2012). Like any sample, of course, these four items cannot represent the full range of combinations of content, context, responses required, and difficulty levels. Nevertheless, it may be useful to consider them here, since they give some specificity to the more general characterisation of numeracy in the survey discussed in the previous subsection. For two of the items, the mathematical contents are framed by Everyday / Personal or Work contexts; for the other two, Society and community contexts⁵. They appear to combine realistic images of the problem at hand with school-like test rubrics, providing the questions that need to be answered, presumably by applying the correct mathematical procedures. Thus these items represent a hybrid type of task.

In any particular country, we need to ask how well these sorts of tasks – such as making precise calculations (as in sample item 3), making precise readings from the appropriate scale (as in item 2), or detecting changes in a time series graph (as 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 4 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 aids⁶. As with any survey, full appreciation of the validity of procedures requires assurance of how these procedures are followed in the field; see OECD (2013b, pp. 47-61). This is even more crucial when results are compared across countries using different fieldwork teams.

³ These levels of difficulty are estimated by Item Response Modelling procedures; see subsection 4.5 below.

⁴ Round 2, including a further 9 countries, is now underway.

⁵ The overall distribution of numeracy items included by contexts was: Everyday / Personal – 45%, Work – 23%, Society – 25% and Further learning – 7% (OECD, 2013b, p.28).

⁶ 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.

External validity

External validity includes the representativeness of the sample for the population of interest; thus, the 5000 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, in order to encourage agreement to be a respondent. Again, judgments about the effectiveness of these procedures depend partly on knowledge of 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 percentages of items correct, is only an *estimate for the corresponding population value* (namely, the mean, or the size of the difference in percentages), for the whole country. Of course, we would like to know about the population value – but this is not possible with certainty, since we only “know” about a subsample. 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 estimate⁷.

Thus, if we consider the PIAAC Numeracy results from OECD (2013a), we would find that the first four countries are:

Japan (288) ... Finland (282) ... Netherlands and Belgium (280)

This appears as a clear ranking – before we realise that a 95% confidence interval for the country score for Finland would be approximately *280 to 284*, and for Netherlands and Belgium, approximately *278 to 282*: thus these countries have *overlapping confidence intervals*, and so their performances are not really able to be differentiated⁸.

Similarly, the differences between the Netherlands and Belgium and the next three ranked countries (the Scandinavians) are not “statistically significant”, again because of the variation that we must always expect in results based only on samples. So what appeared to be a neat ranking of the top 7 dissolves into Japan at the top, followed by a group of six countries, within which one cannot really differentiate performance on the PIACC Numeracy survey (OECD, 2013a, pp. 79-80).

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 can be expected to help greatly with reliability (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 construct an item which can validly assess a respondent’s *reasons for his/her answer*, when the item is computer-marked. 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

⁷ The margin of error depends on the degree of “confidence” desired in the estimate, but is normally *2 standard errors* for a *95% confidence interval*.

⁸ The confidence intervals produced here are only approximate for the sake of illustration: I have estimated the margin of error for country scores based on an inspection of Figure 2.6a (OECD, 2013a, p.80), and have used the idea of countries “with overlapping confidence intervals”, instead of the broadly equivalent idea of countries “differing by an amount which is not statistically significant”.

carry out tasks involving numeracy, and so may not facilitate their “typical” thinking and behaviour responses⁹. Again, similar dilemmas arise for much educational assessment.

Beyond methodology: variations in interpretation and reporting

This discussion of several 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 (Radical Statistics Education Group, 1982 / 2012). 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. These processes require an understanding of the policy context and the ideological debates that surround the reception of results in a particular country, as well as 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. *EERJ* (2012), on the way that PISA results have been reported and used, and in particular, Carvalho on the “plasticity of knowledge” (2012, pp. 180-83). 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 respondents were presented with different sets of items, some “harder”, and some “easier”. So Item Response Modelling is used to (“psychometrically”) estimate a standardised score (e.g. for PIAAC: scores 0-500, mean 250, standard deviation 50). Then, the numerical score is usually related to one of five general “levels” of literacy or numeracy to make it meaningful; see OECD (2013b, pp. 69-70).

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 formerly done. But as in other national and international surveys, there is debate about use of a simple and *one-dimensional* characterisation of an adult’s numeracy. 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 (say, “data and chance”) more difficult than type B items (e.g. “dimension and shape”) – and others find the opposite.

Similarly, some policy-makers attempt to stipulate “the minimum level of numeracy needed to cope with the demands of adult life” in their particular country. But this notion too is questionable, since such generalising claims group together adults with different work, family and social situations, and different literacy / numeracy demands on them; see Black & Yasukawa’s (2013) discussion of current debates in Australia.

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

⁹ And this may disadvantage some groups of respondents more than others, e.g. older ones more than younger. (I am indebted to one of the anonymous referees for this suggestion.)

Some illustrative results for PIAAC from Australia

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

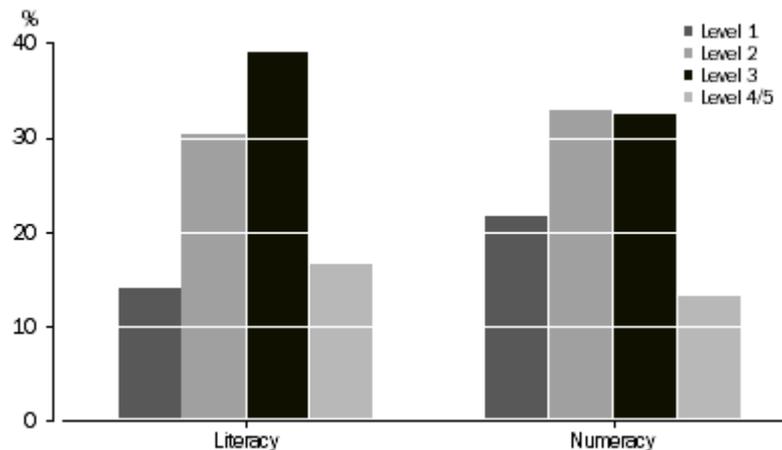


Figure 1. Overall results from PIAAC for Literacy and Numeracy: Australia, 2013

Source: ABS (2013)

Figure 1 allows us to read off the proportions of Australian adults at different skills levels. Approximately 44% (7.3 million) of Australians aged 15 to 74 years had literacy skills at Levels 1 and 2, a further 39% (6.4 million) at Level 3 and 17% (2.7 million) at Levels 4/5. For the numeracy scale, approximately 55% (8.9 million) Australians were assessed at Levels 1 and 2, 32% (5.3 million) at Level 3 and 13% (2.1 million) at Level 4/5. One could also compare literacy and numeracy levels for subgroups, e.g. residents of different Australian states (using other data). For example, the Australian Capital Territory recorded the highest proportion of adults at Level 4/5 (23%) numeracy. We can also ask about gender differences, of interest in much earlier research; see Figure 2.

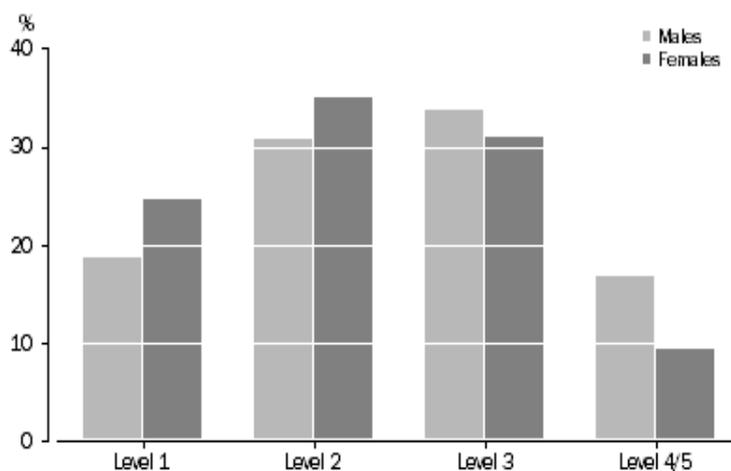


Figure 2. Proportion at each PIAAC numeracy level, by sex: Australia 2013. Source: ABS (2013)

In Figure 2, there appears to be little difference in the proportion of males and females at each level of the numeracy scale. However, a higher proportion of males (17%) attained scores at Levels 4/5, compared with females (9%), as seen from the graph.

We can look at age differences too, over the age group surveyed in Australia: 15-74 (a wider age range than required by PIAAC protocols); see Figure 3.

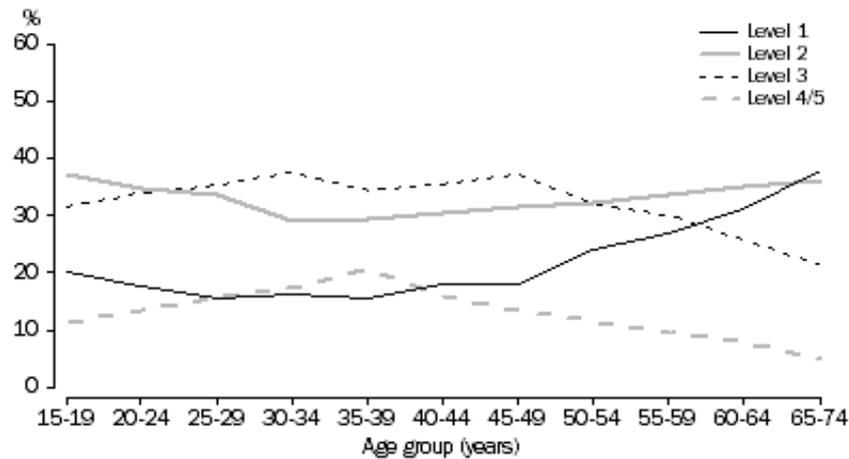


Figure 3. Proportion at each numeracy level, by age: Australia, 2013.

Source: ABS (2013)

The data suggest that proportions of people at Level 1 are highest among the oldest age groups (people aged 60 years and older), and lowest in the younger age and middle-aged groups (people aged 20 to 49 years) for numeracy skills.

Discussion: Possible effects of international surveys and “countervailing forces”

In previous sections we have described the developing role of a globally promoted type of pedagogic discourse promoted by transnational organisations, which asserts adults’ need for certain rather generic skills, and individual countries’ needs to assess these in a comparative way. Basil Bernstein’s analysis (2000) of the structuring of pedagogic institutions and discourses and his focus on changing forms of educational knowledge and practices, along with related work (e.g. Moore with Jones, 2007), can illuminate and critique shifts in the mode of governance of educational policy, in which international surveys like PIAAC are used (by a number of policy actors) to play a role (Tsatsaroni & Evans, 2013).

The international studies of adults, like IALS 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 definition of numeracy (see Section 3 above) and the use of existing classifications of mathematical content. Tsatsaroni and Evans (2013) originally predicted that 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 prescribes what knowledge the adult populations *in all societies* should value, strive to acquire, and demonstrate” (emphasis added). In the event, Christine Pinsent-Johnson’s more recent paper (2013) on adult literacy shows that this “possibility” has already materialised in the Essential Skills in Canada, “a competency-based compendium of employment related ‘learning outcomes’ that integrates [international testing] constructs”. Ontario, Canada’s largest province, has recently begun to use a new curriculum that was put together using these constructs: “A hypothetical and abstracted literacy devised for large-scale testing has been transposed into a

pedagogy that is distinct from schooling and academic literacy practices, and disconnected from personal, community and work literacy practices” (Pinsent-Johnson, 2013, p.2).

There are a number of possible effects of such performance surveys, which may 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 *EERJ*, 2012).

The emerging discourse supported by international surveys may also have effects on teachers’, learners’, researchers’ and citizens’ ways of understanding adult literacy and numeracy¹⁰. Knowledge can come to be seen as generic skills, flowing from a *decontextualised imagining* of the adult’s everyday practices. To the extent that different social groupings and different countries embrace such ideas, they may have restricted access to the countervailing principles of thinking that disciplinary or professional forms of knowledge can provide.

Now, “disciplinary knowledge” can also be understood as “*powerful (mathematical) 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 of a big idea in mathematics / statistics that was illustrated several times at the ALM-20 conference is the idea of *conditional probability*. This idea occurs under many guises: as having the right denominator for your proportions, in arithmetic; 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.) However, for big ideas to be fully appreciated by learners, a coherent curriculum for adults’ mathematics is necessary.

As for positive effects, we must investigate whether international surveys afford opportunities for further research. One can relate performance scores to demographic and attitudinal data from the Background Questionnaire, and/or further information available on numeracy related practices and “use of skills” at work; see OECD (2013a, pp. 101-140) for such analysis, at the international level. These studies 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%) (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 Tsatsaroni & Evans, 2013).

Resources for researching additional interesting questions suggested by the preliminary results are now more accessible than before. OECD makes available, on their website, datasets from PIAAC – and software for data analysis – for research purposes (see <http://www.oecd.org/site/piaac/#d.en.221854>).

In the international adult numeracy community, we can look to alternative research programmes to assert the value of *alternative conceptions of educational knowledge*, and to critique developments in adult educational policy issues, including literacy and numeracy. From within adult numeracy, or what can be called adults’ mathematics education (Evans *et al.*, 2013) – we can illustrate ways to challenge the currently dominant ideas of numeracy and adult skills. For example, Diana Coben and colleagues have challenged the conventional “deficit” characterisation of practising adults’ (nurses’) numeracy, and argued that the high-stakes testing programmes used have often deployed instruments which lacked reliability, validity, and authenticity (Coben, 2000). Hoyles, Noss, Kent & Bakker (2010) go beyond a narrow definition

¹⁰ And lifelong learning more generally (Evans, Wedege & Yasukawa, 2013).

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 & 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 currency conversion to the euro in the Slovak Republic. Gelsa Knijnik and her colleagues (e.g. Knijnik, 2007) 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. The proposals of Knijnik and colleagues and Hoyles et al. are clearly moving towards the formulation of alternative, coherent curricula based on the big ideas that their researches are pointing towards, and helping to develop. Coben and her colleagues are working to develop alternative methods of assessment for professional practitioners.

Powerful knowledges of 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.

To summarise, it seems clear that PIAAC and other international surveys will be key background features in educational policy discussions and educational research for the foreseeable future. These surveys will have a range of effects, some of which will be a focus of struggle involving their transnational sponsors, countries and their citizens. PIAAC itself includes a complex set of measures, and offers the opportunity to relate them in a range of ways. Like all studies, because of its conception and its methodology, it tends to highlight and to emphasise particular aspects of the world it surveys – such as a generic conception of numeracy and literacy, and the use of measures understood as comparable across a globalised world – and to play down others. It is therefore essential for all those interested in adult numeracy and literacy to read its results carefully and sceptically.

Acknowledgements

I thank Anna Tsatsaroni, Tine Wedege and Keiko Yasukawa for useful discussions concerning the arguments in this paper. I also thank other colleagues in ALM, and in particular the audiences at the ALM-20 conference, for stimulating exchanges, and the anonymous referees for careful and helpful comments. Appreciation is due also to colleagues in the PIAAC Numeracy Expert Group for valuable discussions in the course of our work.

References

- ABS (2013). *Programme for the International Assessment of Adult Competencies, Australia, 2011-12: Preliminary Findings*. Retrieved from:
<http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4228.0~2011-2012~Main%20Features~Preliminary%20findings~99>
- Bernstein, B. (2000). *Pedagogy, symbolic control and identity: Theory, research, critique* (revised version). Oxford: Rowan & Littlefield Publishers.
- Black, S. & Yasukawa, K. (2013). Level 3: another single measure of adult literacy and numeracy. *Australian Educational Researcher*. Online:
 DOI 10.1007/s13384-013-0097-9
- Brown, P., Halsey, A. H., Lauder, H. and 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 Society* (pp. 1-44). Oxford: OUP.

- 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>
- 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.
- CSO, Ireland (2013). *Surveys and Methodology: PIAAC*. Online: <http://www.cso.ie/en/surveysandmethodology/education/piaac/%20>
- EERJ (2012). Governing knowledge? PISA in focus. *European Educational Research Journal*, 11(2); *Special Issue* (edited by L. M. Carvalho, S. Grek, E. Mangez & X. Pons). Retrieved from: http://www.wwords.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: RoutledgeFalmer.
- Evans, J., Wedege, T. & Yasukawa, K. (2013). Critical Perspectives on Adults' Mathematics Education; in M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick and F. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 203-242). New York: 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 Topic Study Group 6: Adult and Lifelong Education, 10th International Congress on Mathematical Education (ICME-10). Retrieved from <http://www.icme10.dk>
- 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. (2013). Is it feasible to develop a culturally-sensitive large-scale standardised assessment of literacy skills? Invited paper for Literacy as Numbers conference, University of London Institute of Education, 17 June 2013. Abstract online: <http://www.uea.ac.uk/literacyasnumbers/abstracts>
- 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: Routledge.
- 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 South Bank University, unpublished.
- Moore, R. with Jones, L. (2007). Appropriating Competence: the competency movement, the New Right and the “culture change” project. In R. Moore, *Sociology of knowledge and education* (Ch. 9). 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.
- OECD (2012). *Literacy, Numeracy and Problem Solving in Technology-Rich Environments - Framework for the OECD Survey of Adult Skills*. Paris: OECD Publishing. Online: <http://dx.doi.org/10.1787/9789264128859-en>
- OECD (2013a). *The OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing. Online: <http://dx.doi.org/10.1787/9789264204256-en>
- OECD (2013b). *The Survey of Adult Skills: Reader's Companion*. Paris: OECD Publishing. Online: <http://dx.doi.org/10.1787/9789264204027-en>

- 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.
- Ozga, J. (2009). Governing education through data in England: from regulation to self-evaluation. *Journal of Education Policy*, 24(2), 149-162.
- Pinsent-Johnson, C. (2013). How learning inequality is made to happen: Curricularizing the IALS and classifying adult literacy learners. Invited paper for Literacy as Numbers conference, University of London Institute of Education, 17 June 2013. Abstract online: <http://www.uea.ac.uk/literacyasnumbers/abstracts>
- PIAAC Numeracy Expert Group (2009) [Gal, I. (Chair), Alatorre, S., Close, S., Evans, J., Johansen, L., Maguire, T., Manly, M., Tout, D.], *PIAAC Numeracy Framework*, OECD Education Working Paper no. 35 (24-nov-2009), OECD Publishing. Retrieved from: [http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=edu/wkp\(2009\)14&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=edu/wkp(2009)14&doclanguage=en)
- Radical Statistics Education Group (1982 / 2012). *Reading between the Numbers: a critical guide to educational research*. London: BSSRS Publications. Available at: <http://www.radstats.org.uk/books/ReadingBetweenNumbers.pdf>
- 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* [DOI: 10.1007/s10649-013-9470-x].
- 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.

Appendix. Sample Items from Current Round of PIAAC (2011-2013).

Numeracy – Sample Item 1

This sample item (of difficulty level 3) focuses on the following aspects of the numeracy construct:

Content	<i>Data and chance</i>
Process	<i>Interpret, evaluate</i>
Context	<i>Community and society</i>

Respondents are asked to respond by clicking on one or more of the time periods provided in the left pane on the screen.



Look at the graph about the number of births. Click to answer the question below.

During which period(s) was there a decline in the number of births? Click all that apply.

- 1957 - 1967
- 1967 - 1977
- 1977 - 1987
- 1987 - 1997
- 1997 - 2007

The following graph shows the number of births in the United States from 1957 to 2007. Data are presented every 10 years.



Year	Number of Births
1957	4,300,000
1967	3,570,959
1977	3,326,632
1987	3,809,394
1997	3,880,894
2007	4,315,000

Correct Response: 1957 - 1967 and 1967 - 1977

Numeracy – Sample Item 2

This sample item (of difficulty level 3) focuses on the following aspects of the numeracy construct:

Content	<i>Dimension and shape</i>
Process	<i>Act upon, use (estimate)</i>
Context	<i>Every day or work</i>

Respondents are asked to type in a numerical response based on the graphic provided.

OECD PIAAC

Look at the thermometer. Using the number keys, type your answer to the question below.

What is the temperature shown on the thermometer in degrees Fahrenheit (°F)?

°F



Correct Response: Any value between 77.7 and 78.3

Numeracy – Sample Item 3

This third item (of difficulty level 1) in the set focuses on the following aspects of the numeracy construct:

Content	<i>Dimension and shape</i>
Process	<i>Act upon, use (measure)</i>
Context	<i>Every day or work</i>

Respondents are asked to type in a numerical response based on the graphic provided.

OECD PIAAC

Look at the thermometer. Using the number keys, type your answer to the question below.

If the temperature shown decreases by 30 degrees Celsius, what would the temperature be in degrees Celsius (°C)?

°C



Correct Response: Any value between -4 and -5

Numeracy – Sample Item 4

This sample item (of difficulty level 4) focuses on the following aspects of the numeracy construct:

Content	Quantity and Number
Process	Act upon, use (compute)
Context	Community and society

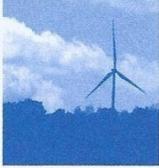


Wind Power Stations

Read the article about wind power stations. Using the number keys, type your answer to the question below.

How many wind power stations would be needed to replace the power generated by the nuclear reactor?

In 2005, the Swedish government closed the last nuclear reactor at the Barsebäck power plant. The reactor had been generating an average energy output of 3,572 GWh of electrical energy per year.



Work continues in Sweden on installing large offshore wind farms using wind power stations. Each wind power station produces about 6,000 MWh of electrical energy per year.

For your information:
 Electrical energy is measured in Watt hours (Wh)

1 kWh	= 1 kilo Wh	=	1,000 Wh
1 MWh	= 1 Mega Wh	=	1,000,000 Wh
1 GWh	= 1 Giga Wh	=	1,000,000,000 Wh

Correct Response: One of the three values (no values between): 595, 596 or 600.

Source: CSO, Ireland (2013)