



National Research and Development Centre  
for adult literacy and numeracy

# **A position paper on international benchmarking**

*(NRDC Project A2)*

**by Greg Brooks, University of Sheffield, and Alison Wolf, Institute of  
Education, University of London**

25 March 2002

## 1. The question

This paper responds to a question posed to the NRDC by the Adult Basic Skills Strategy Unit, namely: How can IALS data cease to be the international benchmark for the level of adult basic skills in Britain? It refers especially to OECD databases and publications since these have been the major vehicle for disseminating IALS-based international comparisons. EU publications, for example, have been less inclined to emphasise IALS or other OECD-sponsored survey results.

The relative attractions of different options will depend on why continuing to use IALS results is seen as unsatisfactory. There are three possible reasons:

1. They are increasingly out of date, since they take no account of the intensive programmes which many OECD countries have undertaken in order to raise basic skills levels
2. They are considered to be inadequate/flawed in some way
3. They do not relate to the National Standards which now form the basis for teaching and learning basic skills in this country, and for evaluating the success of the National Strategy.

We do not make any judgements about the relative importance of these issues, but do distinguish between them in our discussion of options.

## 2. Clarifications

First, we need to point out that the headline IALS (International Adult Literacy Survey) data, as published by the OECD, refer to the whole of the UK. These data were gathered by the Office for National Statistics (ONS) in 1996 (Carey *et al.*, 1997), and within the ONS report separate estimates are given for England, Wales and Scotland. Because the population of England constitutes about 85 per cent of the population of Britain, the estimate for England does not differ markedly from that for Britain as a whole. An enhanced sample for Northern Ireland allowed publication of these results separately.<sup>1</sup> Devolution is likely, in the future, to mean that decisions on whether or not to participate in surveys will often differ within the UK. All the possibilities for future action discussed below can be undertaken by England alone, or by England plus other UK partners.

Secondly, the IALS data cover only literacy (and indirectly an aspect of numeracy). For adult literacy, IALS is so far the only international study ever undertaken.

For adult numeracy there is theoretically more information, but it is not convincing. There have been two international comparative surveys of numeracy or of a placeholder for it. Within IALS, performance on the 'quantitative literacy' domain was highly correlated with performance on the other domains, reflecting the fact that items measured competence in arithmetical calculations where the information needed for the calculations was embedded in text. This reflects a practice common in national and local assessments of (functional) numeracy, which generally do not address

---

<sup>1</sup> The small relative size of Northern Ireland, Scotland and Wales means that separate estimates for these countries generally require a special sampling design, with proportionately larger samples of the population so as to obtain enough cases to allow for separate estimates. Under devolution, the decision on whether to participate in a survey, and fund the enhanced sampling, rests with the individual governments.

competence in decontextualised arithmetic (or in most non-arithmetical aspects of mathematics). However, so far as we know, the IALS quantitative literacy data are not treated as an international benchmark for adult numeracy.

The International Numeracy Survey, also carried out in 1996 and commissioned by the BSA (Basic Skills Agency, 1997), did address decontextualised arithmetic, but had very few items, provided only raw and not internationally standardised scores, did not use probability sampling and has justifiably been ignored for purposes of international comparisons (Coben, 2001). Therefore there is currently no international benchmark for adult numeracy in Britain.

### 3. Main options

So how might IALS data cease to be the international benchmark for adult literacy in England? As far as OECD publications are concerned, in the short term they are likely to remain (alongside the new PISA results for 15-year-olds, also from an OECD sponsored study, and one in which the UK performed well). Attempting to have them deleted, via OECD officials or ministerial meetings, is unlikely to succeed and might even be self-defeating, since it might draw further attention to the data.<sup>2</sup> Of course, eventually the data will seem so out of date that no further credence will be given to them. However, if a more pro-active strategy is desired, then there are effectively only two options:

1. replacing IALS with new international data (i.e. data from a survey which, like IALS, or PISA, or the IEA studies, uses the same questions in a variety of countries)

or

2. replacing IALS with data from national surveys.

We consider these in turn.

#### **Replacing IALS with new international data**

Here we see three short-term options: repeating IALS nationally; taking part in its successor, ALL (Adult Literacy and Lifeskills Survey); or using items from PISA (Programme for International Student Assessment) with an adult population.

##### *Repeating IALS nationally*

Re-running IALS in England as a stand-alone survey would be quite straightforward. If the main reason for concern over the use of the original IALS estimates is that they are now increasingly out-dated (because of the impact of policy changes on attainment), then this has obvious attractions; and new IALS results would probably be used willingly by OECD and other publications.

*However, if dissatisfaction with current use of IALS figures stems from other causes, then these will remain.* One major issue is that the IALS results cannot be converted into results which relate to the National Standards. The ONS development study for

---

<sup>2</sup> As is now well known, when the French government decided to withdraw its data from IALS, because it considered the survey to be biased and the results inaccurate for France, OECD brought considerable pressure to bear in an attempt to reverse this decision.

the proposed National baseline Basic Skills Survey rejected the re-use of IALS items for this reason (Foster, 2000). The conceptual framework used for IALS is not the same as for the National Standards and, as noted above, it covers only a sub-set of adult basic skills.

Any replication of IALS would also, of course, suffer all the defects of the original. IALS has been the subject of a range of criticisms (see, e.g., Sticht, 2001; Blum *et al.*, 2001), some of which emerged from a major re-examination of the methodology and results which was funded by the European Commission (Carey, 2000). Among these criticisms is the lack of differentiation within IALS Level 1 (which is one important reason why IALS cannot be aligned with the National Standards), and the problems caused by the method used to allocate respondents to a level – using different, equally defensible methods produces very different national profiles.<sup>3</sup> Some of the major criticisms of the original international results would not apply to an England-specific re-run of IALS: for example, the very different non-response rates among countries, and the major changes in item difficulty created by translation from English into other languages. However, there would be new ones, especially the aging of some of the items.

### *Taking part in ALL*

The only major international survey currently planned in this area, results from which could replace the IALS ones, is the Adult Literacy and Lifeskills Survey (ALL: earlier known as ILSS, International Life Skills Survey). Appendix 1 lists the countries currently expected to take part in the survey, which will occur this year (2002). ALL is designed/ run by ETS with Stats Canada: OECD supports it though it does not sponsor it formally. A decision was made some while ago that England would not take part in ALL, for reasons discussed below. However, it would still, in principle, be possible for this decision to be reversed, and we therefore review what ALL would, and would not, offer as a possible replacement for IALS.

A feasibility study for UK participation in the study carried out by ONS (Carey and Morris, 1999) concluded that the major potential benefit of ILSS participation was that it would provide the first international estimates for numeracy. The literacy element would effectively re-run IALS, and the methodology for the problem-solving component (now called analytic reasoning) was considered very weak. The numeracy component of ALL appears to be much the best designed, and is, in our view, a vast improvement over both the quantitative numeracy domain of IALS and the International Numeracy Survey. In the original design (i.e. the one extant at the time of the ONS study), countries were required to sign up to all or none of the components: and it seems to have been felt that this was much too expensive for the benefits offered. However, it now appears that the policy has changed, and that England could in principle just carry out the numeracy component. This would provide numeracy estimates based on international tests. Participating in the numeracy tests alone would reduce the costs and the required sample size.

As noted above, one problem with IALS was the lack of differentiation within level 1. A supplementary study has been added to the ALL design which is intended to provide

---

<sup>3</sup> IALS allocates respondents to levels at which they could do 80% of the average items allocated to that level: no credit is given for correct answers to items at higher levels. A recent research note by Sticht (2001) discusses why the 80 per cent criterion is arbitrary, and the effect of setting a different criterion (e.g. 50 per cent).

much greater differentiation within IALS Level 1 literacy – though this increases the cost again.

ALL data could certainly be used to replace IALS data in official publications. However, like IALS, the survey is not aligned with the National Standards and could not be used to report on national attainment in terms of the Standards in either literacy or numeracy. The basic methodology and conceptual framework for literacy remain as for IALS.<sup>4</sup>

### *Using PISA*

This international comparative study of the attainments of 15-year-olds took place in 2000 (OECD, 2000). Within the UK, England, Scotland and Northern Ireland took part, as separate jurisdictions. Wales did not but the results are reported, by OECD, as being for the UK since the absence of Welsh data/sites did not bring the UK results above the minimum exclusion level. The main focus of PISA was reading, with maths and science as minor domains. In contrast to IALS, in PISA England did very well, with low estimates of the proportion of 15-year-olds having literacy problems, and a high position in the 'league tables'.

PISA results will be widely published and publicised. However, the PISA data cannot replace IALS in international databases which deal with adults. Only one age level was studied, from which it is not possible to generalise safely to larger populations; and that age-group falls just below even the widest definition of 'adults'. Nor are the items suitable for use with adults, being designed specifically for a school-based population. As with all other international surveys, the results also cannot be reported in terms of this country's national standards (or, in this case, National Curriculum levels). Performance in reading literacy is presented in terms of five levels (as for IALS); and for 'mathematical literacy' and 'scientific literacy' in terms of mean scores.

PISA offers an alternative set of fully developed items: and in principle, it might seem attractive to use these with adults as an economical alternative to a completely new international survey, and one with some advantages vis-à-vis ALL, or an IALS re-run. However, given that these items were *not* designed for adults, but for 15-year-olds studying curriculum-led maths, science and language, this option also has some obvious and major disadvantages. Moreover, it is unlikely that enough countries would undertake such a survey to produce broad international comparisons and benchmarks.

---

<sup>4</sup> As this discussion indicates, some of the arguments that led to ALL being seen, originally, as too complex and expensive for the likely benefits may now have less force. Greg Brooks has been in touch with one of the principal architects of ALL, Scott Murray of Stats Canada. He made it clear that, despite earlier statements from the co-ordinators of ALL, it is possible for participating national governments to opt out of components of ALL. In particular, analytic reasoning (problem-solving), which many see as its most problematic component, can be omitted. Omitting analytic reasoning reduces the required sample size within any one jurisdiction from 7,000 to 3,250, which in turn considerably reduces the cost. It might still be possible to participate in ALL, since the sponsors would, on the basis on ONS's expertise, be prepared to let England enter the main survey on the basis of only a small pilot. This would have to be completed in time for main data collection to start in January 2003. If there are further data-gathering rounds (IALS had three, in 1994, 1996 and 1998), participation would also become easier. It would be worth keeping a watching brief on ALL in case the opportunity arises to take part in a component, although clearly the government would need to be convinced that the survey design was robust and its content was valid, and that it represented good value for money.

## *The future*

We have discussed these three options at some length because all could, in principle, be undertaken within a reasonably short time period. However, in the medium to longer term, England might wish to encourage and contribute to the development of a new international survey which improves on the methodology of the old. The French government, for example, is currently developing a new adult literacy and numeracy survey as part of a major adult literacy project: this survey is one which on which they would welcome international collaboration in the future. In addition, at a recent meeting in Geneva, the OECD is reported to have looked ahead to the possibility of a new adult survey in 2008.

### **Replacing IALS with national data**

The previous section summarises the current possibilities for replacing IALS estimates with results from other international surveys. This section considers the feasibility of, instead, replacing them with results based on national survey data.

In England there have been to date five purely national surveys of adult literacy and four of adult numeracy, where 'purely' implies 'not part of an international study' (of which in any case there have been only two, IALS and the International Numeracy Survey). The two papers already supplied by Greg Brooks summarise the results of the national surveys. All but two of the purely national surveys were based on samples of adults of particular ages, not on population samples across the full age range from 16 to 60+. (The exceptions are the 1995 survey of writing (Basic Skills Agency, 1996), and the ACACE numeracy survey of 1981 (ACACE, 1982; Social Surveys (Gallup Poll) Ltd, 1981), both of which were based on population samples of people aged 16-60+. However, the survey of writing focused solely on spelling and punctuation, and the ACACE survey did not use probability sampling.) Also, none of the national surveys is more recent than IALS, and two took place over 20 years ago. All pre-date the development of the National Standards.

The issue therefore becomes whether or not it might be possible in the future to use national data to present English adults' literacy and numeracy attainment in a form suitable for international comparisons and inclusion in, e.g., OECD indicator publications. This approach has the advantage of avoiding the problems with bias that so often follow when items are translated into different languages, and also the problem that a single test battery will match different countries' concerns (and curricula) more or less well. Both of these are rightly of concern to politicians as well as assessment experts. However, at a technical level, the answers to whether it is feasible are "Only with great difficulty" and "It depends on how much precision is required."

As we have noted earlier, it is simply impossible to convert scores from a test constructed to measure certain skills, with a given underlying conceptual framework, directly into information about people's (potential or notional) performance on a different set of skills or in relation to a different curriculum. That is why one cannot move from IALS results to conclusions about performance in relation to the Standards, or vice versa. This problem is not a new or unique one: and the most common approach to resolving it is via the use of an **anchor test**.

The idea of an anchor test is that it covers elements common to all the tests or measures that one wants to equate: and that the relationship between each such test

or measure and the anchor test is then established. Anchor tests have been most studied in the United States, because of the desire to equate scores on the different tests used in different states and school systems. (N.B. There is no national set of examinations in the US, and a recent attempt to create national 'voluntary' tests has not succeeded.) In anchor tests developments, the material common to all tests is used to develop an additional test; and samples of students take both the anchor test and the test they normally take (whether a state-set test or one developed by testing agencies such as ETS). The relationship between scores on the anchor test and the other test can then be studied, and one can, in principle, compare scores on one test with scores on another via the anchor test – e.g. by simply converting test scores into their anchor 'equivalents'.

Unfortunately, the consensus of those involved in such developments is that the process does not really work. A major problem, which becomes more serious the more curricula differ, is that the anchor test can only cover the common part of what the baseline tests cover: yet the ambition is to use it to compare and convert scores from complete tests. (Suppose you have two tests: test X covers a, b and c and test Y covers a, d and e. The anchor test can only deal directly with a. In what sense can it be used to say that one person's score on X is the 'same' as another person's on Y given that they have been tested on – and probably learned – quite different things?)

If scores on national tests were to be made comparable with each other, and so used for international comparisons, there would have to be some sort of agreed procedure for conversion: and an anchor test, covering shared material, would be an obvious way forward. In principle it might be easier to identify large amounts of shared content in the field of adult literacy and numeracy than at school level, where there are major differences in curricula between countries. However, identifying that common core and then constructing an anchor test which was felt to be comparably valid after translation into all relevant languages would be a major task, not least at political level. It should also be noted that the final reporting could not, for international comparative purposes, be in terms of England's own national curriculum for basic skills: a shared reporting format tied to the anchor test would have to be agreed and accepted. It also needs emphasising that anchor tests do not and cannot deliver highly precise equating of very different original instruments, and may not remain equally relevant to different curricula over time.

While any enterprise of this type can only be seen as a very long-term objective, some experimental work on this approach has already been undertaken for reading (Bonnet *et al.*, 2000). This is in response to the major defects found in IALS, especially for countries where items were translated from the original English – as noted earlier, a major EU-funded study confirmed the existence of serious bias as a result of translation (Carey, *op. cit.*). The original work on this approach was carried out by England, Finland, France and Italy: Appendix 1 lists all the countries currently participating in the research. Work to date has concentrated on students in secondary school (years 9, 10 and 11).

The results of the research to date indicate that sub-tests with no major evident bias can be used for equating purposes. However, this involves identifying sufficiently large groups of bilingual pupils.<sup>5</sup>

---

<sup>5</sup> Equating must be based on bilingual pupils in order that second-language effects do not confound the results. For example, if a group of native French -speaking pupils who have English as a second language

Anchor test development also confirms that the underlying constructs which countries test in their national tests are quite distinct. For example, the English researchers (from QCA) who participated in the work had to add an additional competence – ‘critical understanding and interpretation of authorial intent’ – to the field originally agreed for the study, in order to align it with the way reading is actually tested in English schools, including in key stage tests. In general, the approach (which is ongoing and now involves eight countries) appears to offer interesting possibilities but no direct and easy way to convert national test scores to a common scale.

#### **4. Concluding arguments**

We have argued that the available international approaches cannot deliver results in terms of the National Standards. We have also argued that an approach based on national data cannot deliver precise equating of scores, or, therefore, a direct replacement for the current international survey approach. However, we have not yet considered international approaches in principle, and we now turn to this aspect.

As we see it, the only way to replace the IALS data as the international benchmark for adult literacy levels in England is to take part in a further international study. Such a study needs to fulfil many requirements of validity and reliability that were not present in IALS or the International Numeracy Survey.

Based on all these considerations, we make the following suggestions.

- 1) IALS should not be re-run nationally.
- 2) In the short to medium term, it is not feasible to replace the IALS data in OECD and comparable reports. (They will eventually, however, be seen by everyone as out of date and no longer relevant.)
- 3) In the longer term, the only source of new, highly specific and internationally comparable data will be a new international study, hopefully one that is better designed than IALS. No such survey, however, will provide results mapped to the National Standards.
- 4) In the short term, ALL should be monitored in case the opportunity arises to take part in a well-designed later round, or part of one.
- 5) And all of this implies that, also in the medium term, England should be working with other countries to build up the international capacity and willingness to mount a well-designed international study. Such a study would certainly cover adult literacy, and possibly also adult numeracy.
- 6) Any decision on future participation by England in such surveys should recognise the severe limitations on what can be expected of the results from such exercises.

---

were to take the test in both French and English, any superiority of performance in French might be due either to greater difficulty in the English version or to the pupils' greater facility in French.

## 5. References

ADVISORY COUNCIL FOR ADULT AND CONTINUING EDUCATION (1982). *Adults' Mathematical Ability and Performance*. Leicester: ACACE.

BASIC SKILLS AGENCY (1996). *Writing Skills: a Survey of How Well People Can Spell and Punctuate*. London: Basic Skills Agency.

BASIC SKILLS AGENCY (1997). *International Numeracy Survey: a Comparison of the Basic Numeracy Skills of Adults 16-60 in Seven Countries*. London: Basic Skills Agency.

BLUM, A., GOLDSTEIN, H. and GUÉRIN-PACE, F. (2001). International Adult Literacy Survey (IALS): an analysis of international comparisons of adult literacy. *Assessment in Education*, **8**, 2, 225-46.

BONNET, G., BRAXMEYER, N., HORNER, S., LAPPALAINEN, H-P., LEVASSEUR, J., NARDI, E., RÉMOND, M., VRIGNAUD, P. and WHITE, J. (2001). *The Use of National Reading Tests for International Comparisons: Ways of Overcoming Cultural Bias*. Paris: Ministère de l'Éducation Nationale.

CAREY, S. (ed.) (2000). *Measuring Adult Literacy: the International Adult Literacy Survey in the European Context*. London: Office for National Statistics.

CAREY, S., LOW, S and HANSBRO, J. (1997). *Adult Literacy in Britain*. London: Office for National Statistics.

CAREY, S. and MORRIS, S. (1999). *International Life skills Survey. UK Feasibility Report*. London: Office for National Statistics.

COBEN, D. (2001). 'Fact, Fiction and Moral Panic: the changing adult numeracy curriculum in England'. In: G.E. FitzSimons, J.O'Donoghue and D. Coben (eds) *Adult and Lifelong Education in Mathematics. Papers from Working Group for Action (WGA6), the Ninth International Congress on Mathematics Education, ICME9*. Melbourne, Australia: Language Australia, in association with Adults Learning Mathematics - A Research Forum (ALM), pp.125-53.

FOSTER, K. (2000). *National Baseline Basic Skills Survey Development Study*. London: Office for National Statistics.

OECD (2001). *Knowledge and Skills for Life: First Results from PISA 2000*. Paris: OECD.

SOCIAL SURVEYS (GALLUP POLL) LTD (1981). *Adult Numeracy Study*. Leicester: ACACE.

STICHT, T. (2001). Has the National Adult Literacy Survey (NALS) Defamed the Competence of America's Labor Force? (Research note).

## APPENDIX 1

Countries participating in:

1. ALL

Argentina

Belgium

Bermuda

Canada

Costa Rica

Italy

Netherlands

Norway

Switzerland

USA

2. Feasibility studies on alignment of national tests

Belgium

England

Finland

France

Italy

Netherlands

Norway

Sweden

## **Acknowledgements**

The authors are grateful to the following international experts who provided information and/or comments on this paper:

John Bynner, London Institute of Education

G rard Bonnet, France

Siobh n Carey, Office for National Statistics

Harvey Goldstein, London Institute of Education

Scott Murray, Statistics Canada

Alistair Pollitt, OCR Examinations Board, Cambridge