

Fall Institute 2011

IALS (INTERNATIONAL ADULT LITERACY SURVEY)  
ITS MEANING AND IMPACT FOR POLICY AND PRACTICE

OCTOBER 23-25, 2011

BANFF, AB

## Groundhog Day?

### The 2009 re-run of the 1996 IALS in Scotland

Author:

Ralf St. Clair

Dean of Graduate Studies

University of Glasgow

Document category:

Country story

The institute offers a forum for diverse opinions and perspectives. The views expressed in this document are the author's, not necessarily those of institute organizers and sponsors.

The author was able to make revisions after the institute and has agreed to post this version.

## **Ralf St. Clair**

*The writer was the Principal Investigator for the Scottish Survey of Adult Literacies 2009 (SSAL2009), building on many years of experience in literacy instruction and measurement in Europe and North America.*

### **Background**

Scotland represents an interesting and unusual case study of the way population level surveys of adult literacy intersect with policy. The main point of the description in this paper is to underline the incompatibility of the IALS instrument with social practices models of literacy. In some contexts this would be unimportant, but Scotland committed to a social practices model of literacy a decade ago (Scottish Executive, 2001), and this has permeated both adult literacy provision and schooling across the nation. Some of the research team, who were involved in designing the social practices-based literacy curricula, found themselves managing the substantial dissonance between the wider philosophical context and what the instrument was designed to do. This dissonance can be considered as a broad threat to the validity of the instrument, as there were real questions about whether the findings would provide information relevant to, compatible with, or usable by, existing policy structures.

Public policy in Scotland is increasingly driven by performance indicators, best represented by the set of 45 key measures in the “Scotland Performs” framework. One of these promises to “Reduce number of working age people with severe literacy and numeracy problems.” The justification for this measure is strikingly economic, referring to people with less developed skills as a “drag” on the economy (Scottish Government, 2007). Given this view of the importance of the basic skills to the economy and the need for accountability, it follows that the creation of a baseline measure against which improvements could be understood was a priority.

In 2008 the Scottish Government decided to survey the population for literacy skills levels by conducting the Scottish Survey of Adult Literacies 2009 (SSAL2009). The last such survey was the IALS in 1996, when Scotland was included as a sub-population of the United Kingdom. After some deliberation, in 2009 the Scottish Government decided to rerun exactly the same instrument as had been used in 1996. The logic behind this choice was that it would provide data that was comparable with 1996 data, but the comparability of both the instrument and the findings was a central problem for those involved in the survey process.

### **The concept of literacy underlying IALS**

To understand IALS and the other tests in the same family such as IALSS, it is important to appreciate the model of literacy embedded in the tests. Any test of anything requires a clear idea of the “thing” being tested. Identifying the construct behind IALS is challenging, as it has only been published in partial accounts that reflect specific phases of its evolution. This reflects the cumulative nature of the instrument’s development, where each stage builds on the previous without restating the fundamental principles of the earlier development. The development process

went backwards and forwards between empirical data and the theoretical framework. It began with a series of test items that were designed on the basis of people's use of text and that varied in anticipated difficulty. In 1985 they were administered to a sample of 3,618 people in the US National Assessment of Educational Progress and ordered in terms of the proportion of the respondents answering correctly. The items were then analyzed in order to identify what sorts of factors made them more or less difficult.

For this set of procedures, task difficulty was defined using the percentage of the population who were able to complete each task successfully. Tasks were systematically represented using a descriptive grammar. Next, two sets of variables were generated from the grammar that appeared to account for task difficulty. (Kirsch & Mosenthal, 1990, p. 25)

In other words, a series of concrete items was tested, and then a model was developed to explain why some were harder than others. The grammar (model) was tested out, and appeared to account for over 80% of the variance in item scores, a remarkably high proportion of the difference in difficulty. This model was applied to the development of the US National Adult Literacy Survey (NALS) and then the IALS instrument. The central assumption of this development process, and every further development of the products, is that literacy ability reflects a universal set of cognitive characteristics that can be reliably assessed through tests of this form.

The statistics involved in the IALS scoring are complex (Kirsch et al., 2000; Sticht, 2001) and have rarely been explained clearly. The decision to use a 500-point scale was part of the original design of the instrument in the early 1980s. The scale represents a trait that can be considered as latent literacy ability, meaning that it is a quality that cannot be directly observed but is assumed to be present to a greater or lesser degree in different people. The aim of the surveys was to estimate the distribution of this trait across the population. Due to the invisibility of the trait, measurement had to be based on probabilities, and the designers decided to use a set of models based on Item Response Theory (IRT) to frame these probabilities. They had a certain number of questions that could be answered correctly or not (there was no partial credit), and had to use these questions to map out abilities across the population. Each question, or item, was assigned a specific level of difficulty on the 500-point scale.

There is a great deal more information available on the details of the process (Kirsch et al., 2001) but there are two key points for this discussion. The first is that the IALS family of surveys approach literacy as a hidden trait that is distributed along a single continuum in the population — this is a necessary assumption for IRT to be effective. The second is that there has to be an assumption that literacy must be cumulative — people who can do 300 point tasks have to be able to do 200 and 100 point tasks as well.

### **The Pragmatics of Process**

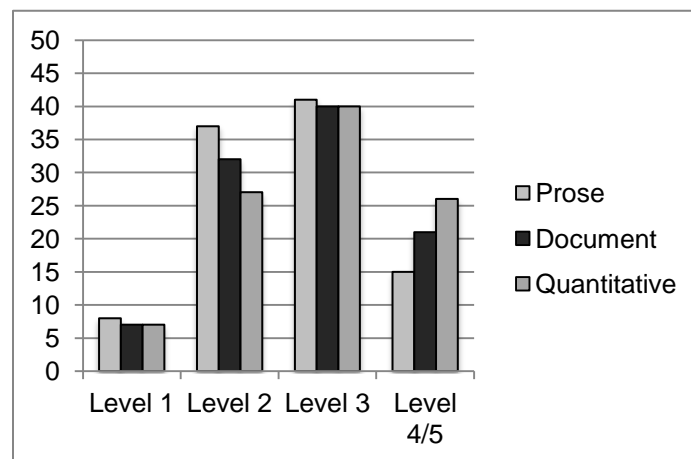
SSAL2009 was implemented by a team of organizations. The main contractors were the Universities of Glasgow and Edinburgh. Fieldwork was conducted by Gallup Europe, based in Brussels, and the data analysis by the National Foundation for Educational Research (NFER),

based in Slough, England. Education Testing Services (ETS) of Princeton, New Jersey, were general consultants and responsible for the scoring and leveling of responses (using proprietary techniques).

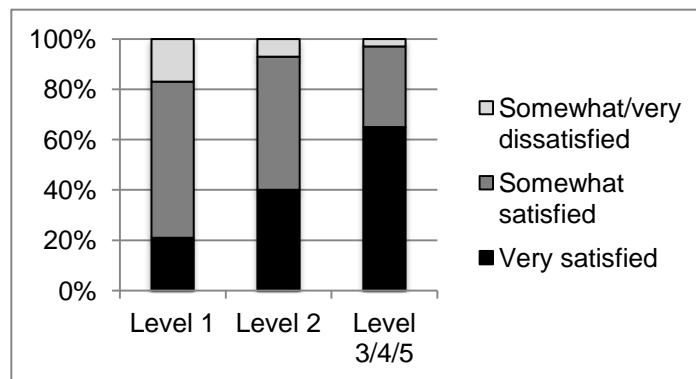
In 1996 the results for Scotland had been based on a sub-sample of the UK survey. This sample was 704 people, limited to the Southern half of Scotland. For the 2009 results to be rigorous, it was important to both expand the sample size and the geographical reach of the survey. In the end, 1953 respondents were sampled from the whole of Scotland (see the technical report for more details, St.Clair, 2010). Respondents responded orally to the background questionnaire with responses recorded by the interviewer, and then used a pen or pencil to complete the screening and cognitive task booklets. The various documents had stickers bearing serial numbers attached to them. The background data was entered online by the interviewers under the same serial number as attached to the hard copy materials. The response booklets came to NFER who scanned them for security and sent them to ETS to be marked. ETS assigned scores and created the five levels, sending the data back to NFER, who linked all the data together to produce the quantitative findings. The universities managed these processes and produced the reports of the findings. The universities were also responsible for making key decisions about the way the data was reported.

### Selected Results

It is useful to look at selected findings of the survey in order to make this discussion more concrete (interested readers are referred to the full report: St.Clair, Tett & Maclachlan, 2010). Figure 1 shows the distribution of the 5 skill levels across the population for each of the three scales. Several aspects of this graph are quite interesting. The first is that there is considerable evidence of strong skills, with good proportions in levels 4 and 5. It is striking that more than 25% of the Scottish population are at level 4/5 for quantitative literacy, for example.



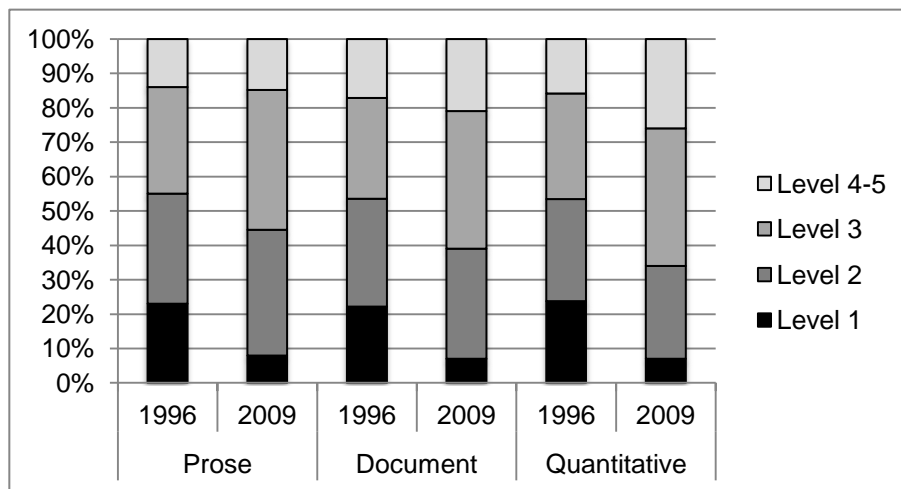
**Figure 1: Distribution of 5 skill levels across the population by scale.** (Source: St.Clair, Tett & Maclachlan, 2010.)



**Figure 2: Satisfaction with skill levels by prose level.** (Source: St.Clair, Tett & Maclachlan, 2010.)

As Sticht (2001) suggests, one unexplained finding of the IALS surveys generally has been the discrepancy between the tested skills of the population, which generally seem to be less strong than expected, and the level of satisfaction people have with their own abilities. Figure 2 shows this for SSAL2009, where over 80% of those who scored at level 1 were somewhat or very satisfied with their abilities. Across the whole population, the proportion was 88%. There are a number of possible explanations for this, but in analyzing the results of SSAL2009 the research team took this finding seriously, avoiding any analysis that reflected a “literacy deficit” or any sense of crisis.

One early expectation of using the same instrument as 1996 was that the 2009 results would be comparable, and Figure 3 shows the results alongside each other. It cannot be stated strongly enough that results were not directly comparable because of different sample sizes, different geographical coverage and an amendment to the leveling model. This graph is included to illustrate the extent that the performance data varied between the two runs of the survey, particularly with regard to the proportions of the population predicted to have skills in levels 1 or 2. This could be interpreted by naïve readers to suggest that there has been a strengthening of skills in Scotland over the last 13 years, but this explanation has no statistical basis.



**Figure 3: Comparison of skill levels, 1996 vs. 2009.** (Source: St.Clair, Tett & Maclachlan, 2010.)

Presentation of the results was a considerable problem for the research team because the results were reported in identical formats but really reflected different statistical approaches. In the end, the authors of the report developed wording that was designed to be unambiguous and represent SSAL2009 as accurately as possible. The proportions facing challenges were based on the population with scores below level 3 on all three scales (26.7%), and those with no scales above level 1 (3.6%). The team concluded:

- Literacy skills in Scotland are comparable with many of the world’s leading economies
- Three-quarters (73.3%) of the Scottish population have a level of skills that has been recognized internationally as appropriate for a contemporary society
- Around one quarter of the Scottish population (26.7%) may face occasional challenges and constrained opportunities due to their skills but will generally cope with their day-to-day lives
- One person in 28 (3.6%) faces serious challenges in their literacies practices
- The proportion of people found to have limited or very limited literacies skills is lower than previous surveys, partly due to better survey methodologies (St.Clair, Tett, & Maclachlan, 2010, p. 2)

This is a different form of presentation from the original OECD reporting, and attempts to recognize both the complexity of literacy and the limits of measurement.

### **Some Thorny Issues**

So far the emphasis has been firmly on the process of SSAL2009. However, making sense of what the findings tell us and how they fit meaningfully into a context requires some deeper consideration. Despite the statistical complexity of the analytical process, the IALS family of surveys is built upon a specific way of looking at specific type of literacy. It is possible to

summarize, somewhat tentatively, the picture of literacy driving the IALS studies. It is seen as based on a set of cognitive attributes, distributed uni-dimensionally across the population, that allows people to read and extract information accurately from instrumental texts and that can be measured in a test situation. This view is quite consistent with the models of literacy circulating in the 1980s but it is a long way from the most influential current views in Europe.

One of the first studies that led to contemporary views of literacy was by Scribner and Cole (1981), who tried to assess the relative contributions of literacy learning and schooling to cognitive development. Their work did not answer that initial question to any extent, as they found all of the concepts to be so deeply culturally embedded that it was impossible to make them meaningful as single constructs. Their findings were reformulated by Street (1984), who argued that literacy must be seen as a social practice because different people interact constantly with different texts in different socially and culturally appropriate ways, and that these varieties were so diverse it made more sense to talk about literacies rather than literacy. The body of work growing out of these early studies, known as the New Literacy Studies, has expanded enormously over the last twenty years and now includes many hundreds of publications. It has also profoundly affected school curricula and research agendas in literacy and language education. The key principle is recognition of the diversity of interaction with spoken and written language, underlining the futility of considering literacy as a single continuum with more able at one end and less able at the other. This also makes concepts such as “literacy deficit” quite useless, since if somebody is unable to complete a specific literacy task, it is simply about a contextualized ability, not an indication of a lack of underlying skill.

It would, of course, be unfair to criticize a survey instrument designed around twenty years ago for failing to take into account more recent theoretical developments, and that is not my intention. The dissonance between the aims of the SSAL2009 instrument and the policy context is worth highlighting, however, as it limited the utility of the findings quite substantially. The central idea of literacy enshrined in any approach to measurement must be compatible with that driving instruction for the information generated to have coherence and application.

The final issue discussed here is comparability more generally. Several decisions were taken in the design of SSAL2009 to improve the quality of the process and the findings. One was the increase in sample size. In 1996, only 704 Scottish residents were involved in the survey, below the sample size required for strong results. In 2009 the research team aimed for 2500 and finally attained 1953 completed surveys. Similarly, in 1996 the sample stopped halfway up Scotland, excluding a lot of rural locations. In 2009 the sample included the whole of Scotland. One further issue is that in 2009 ETS applied a slightly different leveling model that appeared to work better with the data. This model, while compatible with the model used for most international surveys, was significantly different from the model used for the UK in 1996. These three issues make comparison between 1996 and 2009 results irresponsible without an analysis that goes back to the raw data and the way literacy levels were calculated for both of the surveys—a process that has not yet occurred.

ETS expressed concerns about the quality of the data collection procedures, based on information that seemed to suggest interviewers had completed nine surveys in one day and other anomalies. These concerns were taken seriously, and were addressed in two ways. The survey company was asked to warrant the quality of its work, which it did, explaining that the issue had been faulty time recording rather than substantive procedure. In addition the results of the small minority of responses affected were checked for broad compatibility with the bulk of the data, and no apparent bias was detected.

In the end, the policy utility of the survey was limited. It was referred to in the development of the current adult literacies policy, but only as proof of the connection between lower scores and factors associated with poverty (Scottish Government, 2010). Ironically this is one of the most responsible uses of the findings, and the researchers were delighted to see this application of the results. However, policymakers expected SSAL2009 to do more, and it did not fulfil these expectations. It did not fit with the philosophical basis of contemporary Scottish approaches to literacy well enough to have instructional implications, but neither did it provide a direct quantitative comparison with earlier statistics. Scotland will be participating in the upcoming PIAAC survey (the latest iteration of IALS-based surveys) and this is likely to produce results more compatible with future surveys, though it will still not address the mismatch in ideas around what literacy is and how it works.

The first lesson to be drawn from this experience is that data and policy must be coherent. There seems little value in measuring a form of literacy that does not match the forms lying behind the policy at school and adult level. At best this provides a limited picture, but one that media and policymakers may interpret as more telling than it really is. Many jurisdictions use concepts of essential skills to bridge between what can be measured and the broader demands of everyday life, and this is a significant improvement in the use of measurement, though care must still be taken regarding the tendency to assume that measuring one thing tells us everything.

The second lesson, on a more methodological level, is that highly reliable findings from a series of surveys requires commitment to the ongoing evolution of that series—repeating the early iterations leads not to high quality and comparable data, but to a perpetual Groundhog Day of failed aspirations.

## References

- Kirsch, I.S. & Mosenthal, P.B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25(1), 5-30.
- Kirsch, I., Yamamoto, K., Norris, N., Rock, D., Jungeblut, A., O'Reilly, P., Berlin, M., Mohadjer, L., Waksberg, J., Goksel, H., Burke, J., Reiger, S., Green, J., Klein, M., Campbell, A., Jenkins, L., Kolstad, A., Mosenthal, P. & Baldi, S. (2001). *Technical Report and Data File User's Manual For the 1992 National Adult Literacy Survey*, NCES 2001-457. Washington DC: U.S. Department of Education. National Center for Education Statistics.
- Rock, D. (1998) Validity Generalization of the Assessment Across Countries. In Murray, T.S., Kirsch, I.S. & Jenkins, L.B. (Eds.) *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey* (pp. 135-142). Washington DC: Office of Educational Research and Improvement.
- Scottish Executive. (2001). *Adult Literacy and Numeracy in Scotland*. Edinburgh: Scottish Executive.
- Scottish Government. (2007). Reduce number of working age people with severe literacy and numeracy problems. Retrieved June 14, 2009, from <http://www.scotland.gov.uk/About/scotPerforms/indicators/literacyAndNumeracy>
- Scottish Government. (2010). *Adult Literacies in Scotland 2020: Strategic guidance*. Edinburgh: Author.
- St.Clair, R. (Ed.)(2010). *Scottish Survey Of Adult Literacies 2009: Technical Report*. Edinburgh: Scottish Government.
- St.Clair, R., Tett, L., & Maclachlan, K. (2010). *Scottish Survey of Adult Literacies 2009: Report of Findings*. Edinburgh: Scottish Government. Available at: <http://www.scotland.gov.uk/Resource/Doc/158952/0043191.pdf>
- Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Sticht, T.G. (2001). The International Adult Literacy Survey: How well does it represent the literacy abilities of adults? *Canadian Journal for the Study of Adult Education*, 15, 19-36.
- Street, B. V. (1984). *Literacy in theory and practice*. Cambridge: University of Cambridge.