

# The limits of levels: Understanding the International Adult Literacy Surveys (IALS)

Ralf St. Clair

Published online: 20 January 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** The International Adult Literacy Survey (IALS), an initiative of the Organization for Economic Cooperation and Development (OECD), was carried out in the early to mid-1990s across more than 20 countries. It was followed in the early years of the 21st century by the Adult Literacy and Life Skills (ALL) survey and the Programme for the International Assessment of Adult Competencies (PIAAC, currently in data analysis). This article reviews the philosophical basis, theoretical underpinnings and data analysis of the original and subsequent IALS-based surveys. The purpose is to inform users of the survey data of what the surveys can, and cannot, provide. The author argues that the key use of these surveys is providing insights into population-level distribution of one form of literacy, namely a particular kind of text consumption in a developed society. He also points out the challenges regarding the use of the survey series for making international comparisons, for documenting change over time and for representing broad models of literacy. The tendency to use the survey findings for these uses is considered by the author as a misuse of the data, which leaves the potential of the IALS surveys to provide insights into the effectiveness and equity of different educational systems largely untapped.

**Keywords** Literacy · Surveys · International · Comparative · IALS

**Résumé** Les limites des niveaux d'apprentissage : comprendre l'Enquête internationale sur la littératie des adultes (EILA/IALS) – L'Enquête internationale sur la littératie des adultes, une initiative de l'Organisation de Coopération et de Développement économiques (OCDE), a été réalisée du début au milieu des années 1990 dans plus de 20 pays. Elle a été suivie au début du XXIe siècle par l'Enquête

---

R. St. Clair (✉)  
Department of Integrated Studies in Education, Faculty of Education, McGill University,  
3700 Rue McTavish, Montreal, QC H3A 1Y2, Canada  
e-mail: ralf.stclair@mcgill.ca

sur la littératie et les compétences des adultes (ELCA/ALL) et par le Programme pour l'évaluation internationale des compétences des adultes (PEICA/PIAAC, actuellement en phase d'analyse des données). L'auteur de l'article dégage le fondement philosophique, les bases théoriques et l'analyse des données de l'enquête d'origine EILA ainsi que des études qui lui ont succédé. Son intention est d'informer les utilisateurs des données de ces enquêtes sur ce qu'elles peuvent et ne peuvent pas fournir. Il affirme que l'intérêt essentiel de ces enquêtes réside dans des renseignements sur la répartition parmi les populations d'une seule forme de littératie, à savoir l'exploitation spécifique d'un texte dans une société industrialisée. Il signale en outre les problèmes liés à l'utilisation de la série d'enquêtes en vue d'établir des comparaisons internationales, de documenter l'évolution dans le temps, et de présenter des modèles généraux d'alphabétisation. L'auteur considère la tendance à exploiter les résultats de ces enquêtes dans ces buts comme un usage abusif des données, qui laisse largement inexploité leur potentiel d'apporter des éclaircissements sur l'efficacité et l'équité des différents systèmes éducatifs.

**Zusammenfassung** Niveaustufen sind nicht alles: Was sagen die internationalen Untersuchungen der Lese- und Schreibfähigkeit von Erwachsenen (IALS) aus? – Die International Adult Literacy Survey (IALS, internationale Untersuchung der Lese- und Schreibfähigkeit von Erwachsenen), eine Initiative der Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung (OECD) wurde Anfang bis Mitte der 1990er-Jahre in mehr als 20 Ländern durchgeführt. Als Nächstes folgten in den ersten Jahren des 21. Jahrhunderts die Adult Literacy and Life Skills Survey (ALL-Studie über die Lese- und Schreibfähigkeiten und Lebenskompetenzen Erwachsener) und das Programme for the International Assessment of Adult Competencies (PIAAC, internationale Vergleichsstudie der Kompetenzen Erwachsener, deren Daten derzeit noch analysiert werden). Gegenstand des vorliegenden Artikels sind die philosophischen Grundlagen, der theoretische Unterbau und die Datenanalysen der ursprünglichen und der nachfolgenden IALS-Studien. Er liefert den Nutzern der Erhebungsdaten Informationen darüber, was die Untersuchungen aussagen und was nicht. Der Autor vertritt die These, dass der wichtigste Nutzen dieser Studien darin besteht, dass sie Einsichten in die Verbreitung einer einzigen Form der Lese- und Schreibfähigkeit, nämlich einer ganz bestimmten Art und Weise der Textaneignung in einer entwickelten Gesellschaft, auf Bevölkerungsebene vermitteln. Außerdem werden die Schwierigkeiten beleuchtet, die sich durch die Nutzung der Erhebungen für internationale Vergleiche, für die Dokumentation des Wandels im Lauf der Zeit und für die Darstellung umfassender Modelle der Lese- und Schreibfertigkeit ergeben. Die Tendenz, die Ergebnisse der Studien in dieser Weise zu nutzen, ist nach Auffassung des Autors ein Datenmissbrauch, der dazu führt, dass das eigentliche Potenzial der IALS-Erhebungen, Einsichten in die Wirksamkeit und Gerechtigkeit verschiedener Bildungssysteme zu gewähren, im Großen und Ganzen unerschlossen bleibt.

**Resumen** Los límites de los niveles: entender las Encuestas Internacionales de Alfabetización de Adultos (IALS) – La Encuesta Internacional de Alfabetización de Adultos (IALS), una iniciativa de la Organización de Cooperación y Desarrollo

Económico (OCDE), fue realizada entre principios y mediados de los 1990 en más de 20 países. En los primeros años del siglo XXI fue sucedida por la Encuesta de Alfabetización y Destrezas para la Vida del Adulto (ALL) y el Programa para la Evaluación Internacional de las Competencias de los Adultos (PIAAC, actualmente en etapa de análisis de datos). Este artículo revisa la base filosófica, los fundamentos teóricos y el análisis de datos de la encuesta original IALS y de las que le siguieron, con la finalidad de informar a los usuarios de los datos recabados sobre qué es lo que pueden y qué es lo que no pueden ofrecer estos estudios. El autor sostiene que el uso esencial de estas encuestas es el de proveer una idea de la distribución de los niveles de conocimiento en la población en cuanto a una forma de alfabetismo; o sea, de un tipo determinado de consumo de texto en una sociedad desarrollada. También destaca los retos que se presentan al usar las series de encuestas para realizar comparaciones internacionales, para documentar el cambio a lo largo del tiempo y para representar modelos generales de alfabetización. El autor considera que la tendencia a usar los resultados de estas encuestas para estos fines es un mal uso de los datos, porque se desaprovecha en gran medida el potencial que ofrecen las encuestas IALS para proporcionar conocimientos sobre la efectividad y la equidad de diferentes sistemas educativos.

**Резюме** Ограничения уровней: Понимание применения международных исследований грамотности взрослых (МИГВ) – Международное исследование грамотности взрослых (МИГВ), инициатива Организации экономического сотрудничества и развития (ОЭСР), было проведено в начале-середине 1990-х годов в более чем 20 странах. В начале 21 века за ним последовало исследование уровня грамотности и навыков жизнедеятельности взрослых (ALL) и программа международной оценки компетенций взрослых (PIAAC, в настоящее время находится на этапе анализа данных). Эта статья рассматривает философские основы, теоретические обоснования и анализ данных первоначального и последующих исследований МИГВ. Цель – проинформировать пользователей данных исследований о том, какая информация может быть почерпнута из исследований, а какая нет. Автор считает, что эти исследования в основном полезны для понимания уровня распространения среди населения только одной из форм грамотности, а именно определенного способа использования текста в развитом обществе. Автор также указывает на проблемы применения серийных исследований для сравнения результатов между странами, документирования изменений на протяжении нескольких лет и их использования для широких моделей грамотности. Тенденция к использованию результатов исследований для этих целей представляется автору ошибочной, что не позволяет в полной мере раскрыть потенциал исследований МИГВ для понимания эффективности и справедливости различных образовательных систем.

It seems that the requirements for measurement are becoming more pressing across the broad field of education, and that the data created by these measurements are having increasingly profound effects upon policies within the field. One example of this trend is the considerable impact that the International Adult Literacy Survey (IALS) of the mid-1990s has had on literacy policy and adult education policy more generally. The survey results suggested that there were large proportions of the population in the most developed countries in the world who were facing serious challenges due to their limited literacy skills. There is evidence that the governments in the UK and Canada, and to a lesser extent in the United States, responded directly and positively to the surveys (St. Clair and Belzer 2007), moving adult literacy education into a more central position in educational policymaking. Equally strong negative reactions were also apparent in some countries, such as in France, where the IALS methods were challenged. This resulted in France running its own survey, which found considerably fewer people living with literacy challenges (Blum et al. 2001). However, the consistent theme across all the different national experiences is that population-level measures of literacy are being taken seriously.

There have been three iterations of international adult literacy studies to date: (1) the original International Adult Literacy Survey (IALS) conducted in 21 countries between 1993 and 1998; (2) the Adult Literacy and Life Skills Survey (ALL) conducted in the early years of the 21st century; and (3) the Programme for the International Assessment of Adult Competencies (PIAAC, currently in data analysis). Across the series, these surveys have become more sophisticated and more capable of capturing information that is of value to educators and policymakers, but it is important to acknowledge that they share an underlying logic and many of the same strengths and weaknesses. The aim of this article is to provide an understanding of these surveys and to support the ability of practitioners and policymakers to judge which aspects of the survey findings are more, or less, useful to their work. The overall argument of this article is that the results of these surveys must be approached with informed caution – they cannot be viewed as unproblematic measurements of literacy skills. Throughout this discussion, “IALS surveys” is used to refer to the entire series of international surveys with a common heritage.

## **Background to the family of surveys**

The IALS was an initiative of the Organization for Economic Cooperation and Development (OECD) in the early to mid-1990s across 21 countries (OECD 2000). It was primarily an attempt to measure the human capital within economies; human capital being the knowledge and skills of the population in a country, which is sometimes measured by educational levels and sometimes by direct assessment of specific abilities (Becker 1975). IALS surveys rest upon the claim that it is possible to understand the amount and distribution of human capital within an economy by understanding the levels and patterns of actual literacy skills. This view entails two important claims. The first is that literacy is at the centre of knowledge within a

developed economy; in essence, it is an irreplaceable gateway to all other types of knowing. If it were possible to substitute other information strategies for literacy, then measuring literacy would tell us very little about human capital. The second is that the focus of surveys should be actual abilities measured by tests and not the levels of education people complete. It is important to acknowledge that neither of these claims has been fully justified by empirical work at any point, and that significant research remains to be done in both of these areas.

The original survey placed people in five levels of performance numbered 1–5, with 5 denoting high performance. The proportions in levels 4 and 5 were always reported as a combined figure, because the numbers in level 5 were too low to be reliable. The surveys claimed to measure three different forms of literacy: prose, document and quantitative. Prose literacy indicates interaction with continuous text (like a newspaper article), while document literacy refers to interaction with non-continuous text, possibly presented in a matrix format (like a bus timetable). Quantitative literacy refers to the ability to extract and apply numerical information from text (for example quantities from a table) and not the kind of operations usually considered as numeracy or mathematics. It was claimed that including this three-dimensional model would allow the survey to capture literacy in a more nuanced way than a single scale (Kirsch et al. 2001), because each scale represented something different and irreducible. While there are indications that the three scales do capture some difference in the forms of literacy, there is very little evidence that they represent fundamentally different dimensions (Rock 1998). In practice, public discussions of the results have tended to consider prose literacy levels alone.

The results of the original IALS survey were unexpected in two ways. The first was the much lower than expected level of measured literacy abilities within each country. In the UK, for example, when the sample results on the prose scale were extrapolated to the population, they suggested that 22 per cent of UK residents were at level 1, 30 per cent at level 2, 31 per cent at level 3 and 17 per cent at level 4/5. Finding half of the population in a developed economy in the lowest two levels after well over a century of compulsory education was a source of considerable concern.

The reports of the outcomes sometimes used emotive language, such as when they referred to people having “a severe literacy deficit” (OECD 2000, p. xiii). This is a highly problematic phrase; not only was education as a field moving away from deficit language by the early 1990s, literacy theory had also demonstrated the lack of utility or rigour underlying such perspectives (cf. Barton 1995). Use of such terminology could give the impression that the results of the survey had uncovered a vast and significant educational and social policy problem. Concerns about the outcomes were only reinforced when the OECD stated:

Level 3 is considered a suitable minimum for coping with the demands of everyday life and work in a complex, advanced society. It denotes roughly the skill level required for successful secondary school completion and college entry (OECD 2009, p. 6).

On this basis, 52 per cent of the UK population did not have the “suitable minimum” for everyday life, and the UK was roughly middle of the range for prose literacy. The US did marginally better, but even the strongest-performing country on

prose literacy, Sweden, had over one in five of its population deemed as below the suitable minimum. At the other end of the scale, more than four out of five people in Chile were claimed not to meet the minimum. Given the concerns about the quality of education systems over the previous 30 years, these were explosive results that demanded an immediate political response.

In retrospect, however, the level 3 claim was perhaps less robust than was understood at the time. There is little apparent empirical basis for considering that level 3 represents the functional literacy level, that it was the same in all participating countries, or that it was equivalent to completing secondary education. Interestingly, the text quoted above was removed from the OECD website during 2010, but it was influential for many years.

The second surprising dimension of the IALS results was the pattern of differences and commonalities between countries when they were compared and ranked by the percentages in different levels. The exact order of countries varied slightly, depending on which of the three dimensions was considered, but generally it was highly consistent and somewhat predictable. Scandinavian countries were at the top, with France and Poland tied at the bottom for highest proportion of their population at level 1. This result was not accepted by France, which withdrew before the results were published. Nonetheless, the OECD promoted the rankings as reliable and useful data.

Accepting the IALS rankings as useful and accurate requires confidence that the same things were tested in the same way in each country. A retest survey was commissioned by the European Union in 1998 and conducted by a team lead by Siobhan Carey of the Office of National Statistics in the United Kingdom (Carey 2000). The findings pointed to some significant issues regarding the way the tests worked in different contexts. One example was that the proportion of correct answers could be expected to vary between the French and Swiss-French population because of the way the instruments were translated into each language irrespective of the skill levels of the sample using the two forms of French (Carey 2000). For this reason, alongside a variety of others, the retest survey suggested that comparisons and rankings between nations should be avoided. Analyses conducted by members of the IALS team had come to the same conclusion several years earlier (Kalton et al. 1998; Rock 1998). Readers interested in the details of these discussions are encouraged to follow up these documents.

The cautions of these analysts contradict the common understanding of IALS, which is often seen among the policy community as a robust comparative measure of skills and human capital (Scottish Executive 2001) and indeed is promoted as such by OECD. Given the use made of the IALS survey findings, it is important to be clear why using them comparatively is a problem. The best way to understand this issue is to look at the design of the surveys in more detail.

## Understanding the instrument

The instrument used for IALS surveys is unique, but based closely upon a set of concepts and measurement approaches that have been developed by Irwin S. Kirsch

and colleagues since the early 1980s (cf. Kirsch and Guthrie 1980). They were first applied in a series of tests such as the US National Assessment of Educational Progress (and the subset that came to be referred to as the US Young Adult Literacy Assessment). These surveys were originally conducted by a company called Westat and the Educational Testing Service, and continue to this day in the work of the National Center for Education Statistics, part of the US federal Department of Education. The key idea from the original conceptual framework, and pervading the descendent tests, is that literacy is at heart a process of retrieving information from a text and using it to complete a task successfully. As the steps needed to complete this process become more complex, the task is considered more difficult.

The survey instrument consists of three sections. The first is a background questionnaire designed to gather a great deal of information on the respondent, such as work and educational history, language use, children, marital status, age, national background, health conditions, living conditions and so on. It takes considerable time to complete this section of the questionnaire. There is then a short screening section with six questions. Only if respondents answer two or more of these correctly can they move on to the main literacy survey. The aim of the screening section is to provide a way to identify people with extremely low skills levels and stop the survey at that point. In fact, very few people in each national survey have ever been filtered by the screen.

The third part of the survey, and the main interest as regards literacy measurement, is what is referred to as the “cognitive” testing. This term is used to indicate that this section is intended to measure the cognitive skills assumed to lie at the heart of literacy abilities, rather than in the more formal psychological sense. In practice, “cognitive” could be replaced with “reading” with no loss of meaning. The questions themselves are very highly protected by copyright and cannot be reproduced here, but Mary Hamilton and David Barton (2000) have conducted an insightful and detailed discussion of the nature of the questions included in the survey.

Each respondent has to answer around 40 questions, allocated on a balanced incomplete block design. This means that there are seven different survey booklets, each containing three blocks of around a dozen questions. The blocks rotate through the booklets, so that booklet A has (for example) blocks 1, 2 and 6, booklet B has blocks 2, 3 and 7, and so on. Because the booklets overlap and each block is completed by 43 per cent of the respondents ( $3/7$ ), this design allows the probable distribution of correct answers to any one block to be deduced from the existing answers, effectively increasing the sample power. The principle is like having two people and giving each two out of three texts to read. If one of the texts is given to both, the researcher can use the performance on the shared text to work out how each would have probably done on the text they did not read. This approach is not very robust for the measurement of the skills of each individual, but it works well for a large number of people. This multiplying up is a useful way to achieve the aim of the IALS surveys of looking at the probable pattern of skills across a population if the underlying – and necessary – model of literacy is accepted.

## Constructing literacy

Research surveys such as the IALS are driven by an underlying construct about the “thing” being measured. Understanding the construct of literacy embedded in the IALS tests requires clarification of the design principles and history of the instrument. Reconstructing that history is challenging, as it has only been published in partial accounts that cover 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 following account is based on a study of the available accounts of the development process, and for brevity I have only included the document literacy scale, though a very similar procedure was followed for the other two scales (Kirsch and Mosenthal 1990).

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 analysed in order to identify what sorts of factors made them more or less difficult. The analysis identified 13 factors, which fell into three groups: the complexity and structure of the document, the nature of the tasks respondents were asked to complete, and the nature of the processes required (ibid.) These groupings were organised into a model referred to as a relational grammar. The test designers sum this up:

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. (ibid., 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 was tested out, and appeared to account for over 80 per cent of the variance in item scores, a remarkably high proportion of the difference in difficulty. Overall, this process

... establishes predictable difficulty levels and known cognitive characteristics for one set of assessment exercises. With this knowledge, we could begin to develop new assessment exercises that systematically manipulate materials, tasks and processes. If successful, such an effort would result in greater efficiency in designing tests and assessments as well as in interpreting particular proficiency levels. (ibid., p. 25)

This approach was applied to the development of the 1992 US National Adult Literacy Survey and then the IALS instrument, as well as later large-scale surveys. The underlying concept is that literacy ability reflects a universal set of cognitive characteristics that can be reliably assessed through paper and pencil tests. This is a very particular literacy construct, reflecting the conviction that reading and writing



can be meaningfully considered as individual matters of mental processing rather than social activities. While there certainly are mental processes involved in using texts, it is quite misleading to reduce literacy to only those activities which can be measured in simple written tests. Apart from anything else, the mental processes of reading are not yet well understood, and considering them as cognitive characteristics that can be reliably and validly assessed on a one dimensional scale (from 0 to 500 in this case) is not justified (St. Clair 2010). This point will recur later in the paper when contemporary models of literacy are discussed.

## Scoring

Many people familiar with regular classroom tests would assume that an individual's performance on the test items would lead to their score. One question might be worth 20 points, another 30, and the person's score would be the total points of the questions they answered correctly. In the IALS family of tests, the approach is considerably more complex, for good reason. The surveys were never designed to measure the literacy of individuals, but to estimate the skills within a population. It is crucial to appreciate that the scores for individuals are effectively meaningless, and function only as a data source for the calculation of population skills levels.

The statistics involved in the IALS scoring are complex (Kolstad 1996; Sticht 2001; Kirsch and Mosenthal 1990) and have rarely been explained clearly. As part of the original design of the instrument in the early 1980s, a decision was taken to use a 500-point scale with a mean score of 250 and a standard deviation of 50 points – meaning that two-thirds of scores would be between 200 and 300 (Kirsch et al. 2001). The scale is intended to represent a latent literacy trait, meaning an intrinsic level of literacy ability 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 (in the three forms of literacy) across the population. When the trait is invisible and cannot be measured directly, measurement has to be based on probabilities. In the case of the IALS instruments, the designers decided to use a set of models based on Item Response Theory (IRT) to frame these probabilities. IRT works by assuming that whether one question is answered correctly can predict whether other questions would be answered correctly because of the common underlying trait. IALS surveys had a certain number of questions that could be answered correctly or not, 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, and could be answered correctly or incorrectly (there was no partial credit). For this measurement approach to provide useful information about population skills, there has to be an assumption that the invisible literacy abilities are distributed across the population in the form of a normal curve.

Probability is a central concept in these surveys. The analysis leverages a relatively small number of responses from a relatively small number of people into statements about an entire population. It does this by saying that a certain

percentage of the population is likely to have a certain level of ability. The likelihood used in the IALS surveys is 80 per cent, so only the percentage of people who are 80 per cent likely to score in level 4 prose literacy (for example) are assigned to that level. In order for this entire system to work, the difficulty of test items has to be reliably linear – there has to be a smaller number of participants likely to score at the 325-point level than at the 150-point level. The background information from the survey was used to strengthen the estimates of probability, so that, for example, middle-aged white men were assigned a different weighting for the estimates than white male teenagers. By making a great many deductions using a sophisticated mixture of actual responses and background information, IALS could make powerful estimates of the distribution of an invisible factor across a population.

There is a great deal more information available on the details of the process (Kirsch et al. 2001), but there are two key points that will recur in this discussion. The first is that the IALS family of surveys approaches 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 there are more people with lower levels of the underlying trait than with higher levels, or else the information derived from a series of correct/incorrect questions would not be helpful. In other words, literacy must be cumulative – people who can do 300-point tasks have to be able to do 200- and 100-point tasks as well, but the opposite cannot be true. These assumptions are completely compatible with the underlying concept of literacy as discussed earlier.

## Levelling

One aspect of the IALS programme that has made it more attractive to policymakers is the division of literacy abilities into five levels, and it is instructive to examine how this process works. The levels came quite late in the evolution of the tests, and were primarily developed to ease communication about the results. The levelling is overlaid on top of the 500-point scale in a simple arithmetical way, with level 1 representing scores up to 225, level 2 from 225.001 to 275, level 3 from 275.001 to 325, and levels 4 and 5 including all scores above 325.001 (since level 4 and 5 are always reported together, the division between them is not relevant). For all three types of literacy, a score of 348 is level 4, and a score of 186 is level 1, etc. It is worth noting that most of the discrimination in the scale has to be between 225.001 and 325, since there are two very important levels (2 and 3) tied to these scores.

Table 1 provides details of the types of questions assigned to each of the five levels, showing the increasing complexity of the tasks. It also demonstrates the consistency of the types of abilities represented by the tasks.

When reading the IALS results, it is important to bear in mind that the model of measurement expects two-thirds of the population to fall between 200 and 300 points. That is, a perfect population score (if such an idea made sense) would have a mean in the centre of the level 2 range (250). This is an extremely important point for the interpretation of the results, even if literacy abilities were genuinely arranged

**Table 1** Description of the five IALS prose levels

Prose	
Level 1 (0–225)	Most of the tasks at this level require the reader to locate one piece of information in the text that is identical or synonymous to the information given in the directive. If a plausible incorrect answer is present in the text, it tends not to be near the correct information.
Level 2 (226–275)	Tasks at this level tend to require the reader to locate one or more pieces of information in the text, but several distractors may be present, or low-level inferences may be required. Tasks at this level also begin to ask readers to integrate two or more pieces of information, or to compare and contrast information.
Level 3 (276–325)	Tasks at this level tend to direct readers to search texts to match information that require low-level inferences or that meet specified conditions. Sometimes the reader is required to identify several pieces of information that are located in different sentences or paragraphs rather than in a single sentence. Readers may also be asked to integrate or to compare and contrast information across paragraphs or sections of text.
Level 4 (326–375)	These tasks require readers to perform multiple-feature matching or to provide several responses where the requested information must be identified through text-based inferences. Tasks at this level may also require the reader to integrate or contrast pieces of information, sometimes presented in relatively lengthy texts. Typically, these texts contain more distracting information and the information that is requested is more abstract.
Level 5 (376–500)	Some tasks at this level require the reader to search for information in dense text that contains a number of plausible distractors. Some require readers to make high-level inferences or use specialised knowledge.

Source Statistics Canada (2008)

along a unidimensional normal curve. For the IRT model to make sense, 50 per cent of the population should score below 250 points or so. Therefore the model is designed on the premise that at least half the respondents will be below level 3. This raises the question of how surprising it really is that around half the population in almost every country was found to lie below level 3, and to what extent this finding is driven by the analytical model.

The IALS instrument is a complex and highly sophisticated approach to understanding the distribution of literacy abilities across a society (for more discussion, see Blum et al. 2001). It uses individual responses to items to estimate the probable patterns of a latent trait, and then turns these estimations into a system of five levels representing capabilities across a population. Having covered some of the technical complexities of the IALS surveys, in the following section I discuss the model of literacy that underpins the survey design and the meaning of the findings.

## Reading the results

This discussion has been quite technical so far, but the critical question for educators and policymakers is what it all means – that is, what can the results of this series of surveys tell us? This section addresses that question, looking at what is

missing from the surveys, how the model of literacy underpinning the surveys fits current theories of literacy, and at the degree of consistency of the survey results over time.

Despite the statistically complex mechanism of the IALS surveys, they are built upon a relatively simple model of literacy ability. One of the key assumptions of this model is that measuring the ability of an individual to locate and integrate information from the text and diagrams presented to them provides information on their overall literacy. As mentioned earlier, the individual scores are not used directly in generating the survey results, so the issue here is not whether the sample relates accurately to an individual's real world abilities, but what types of literacy abilities are considered in the overall model. No research can include everything, and the IALS surveys have chosen specific forms of interaction with text to act as proxies for literacy more broadly; inevitably, this means that some forms are left out.

One of the most interesting examples of what is not measured is writing. IALS provides no information whatsoever on writing (beyond a few questions asking whether people are happy with their own writing), meaning that a very important component of literacy is not considered by the survey. This is the result of a deliberate decision by the original test designers, who adopted a definition of literacy as "using printed and written information to function in society, to achieve one's goals and to develop one's knowledge and potential" (Kirsch and Jungeblut 1986, p. 3). *Producing* printed and written text is not emphasised in this definition. Interestingly, the first study in the series leading up to the IALS surveys did include some writing tasks (Kirsch and Jungeblut 1986), but they seem to have been dropped by the time of the IALS. The irony of not discussing writing is that the IALS surveys require respondents to write in their answers, making writing ability a central concern of the tests – albeit a silent one.

While the definition of literacy used by the IALS is about reading, that is *consumption* of text, there are further limits to the types of texts. There is, for example, no poetry to interpret. The texts provided in the test booklets are extremely instrumental and direct in their format. It seems reasonable to deduce from this exclusion that the measure does not extend to literary uses of language, though this has never been addressed directly. At the same time, a high number of the texts have to do with shopping and being a consumer, with money management or with employment. Thus the surveys make an assumption regarding the types of text that are most functional for adults in OECD countries which is not based on any data or even discussed at any length in survey documents. None of the texts are critical or political, and participants are not asked to respond to arguments. The surveys seem to centre on a highly limited set of texts that reflect a highly selective set of text consumption in a developed society.

Another limit is that the texts are provided in the standard language for the country in which the surveys are administered. No account is taken of dialect or local varieties of written language (again privileging a specific type of text), thus people who do not read and write that standard language are excluded from the survey. Depending on national context, this could be a significant issue. For countries with high rates of immigration it is likely that some of the most educated

and literate members of the population would be left out of the survey. There is no way of knowing with certainty whether this is a significant source of bias in the results, but once again it provides clues about the model of literacy being implemented (cf. Hamilton and Barton 2000).

Finally, the surveys are a test and in many ways they are similar to school tests, despite the genuine efforts to use “real-life” examples of texts. As with any test, there is a conflation of the ability to perform the task with the ability to take tests, especially under observation by a stranger. It is not possible to know the difference this element made, though one study that retested original respondents in different circumstances showed a remarkable increase in test scores (Carey 2000). The lack of discussion of this issue in the documentation indicates that the test designers did not consider comfort with tests to be a significant factor, even though many of the adult respondents were likely not to have taken a test for 40 or 50 years, and may well have had unhappy memories of the last time. It could be suggested that the comfort with testing procedures assumed for the original respondents (American young people in the early 1980s) was taken for granted in all subsequent iterations of the test format.

Overall, it is possible to summarise the picture of literacy driving the IALS studies. It is based on a set of cognitive attributes that allow people to extract information accurately from instrumental texts and that can be measured in a test situation. For educators and policymakers who require information on this form of literacy, the IALS surveys are a high-quality source of data, but it is important to be aware that the results of the surveys do not go very far beyond this model, and certainly do not represent a complete measure of human capital or literacy abilities.

Many literacy policymakers and educators are interested in a broader view of literacy, and this view evolves along with theories of literacy and the current social challenges. Perspectives on literacy have changed very significantly since the beginning of the development of the IALS series in the 1980s. One early study contributing to these new perspectives was conducted by Sylvia Scribner and Michael Cole (1981), who tried to assess the relative contributions of literacy learning and schooling to cognitive development. Their work did not manage to answer that initial question to any extent, as they found all the concepts surrounding text use to be so deeply embedded in culture that it was impossible to make them meaningful as single constructs. Their findings were reformulated and extended by Brian V. Street (1984), who argued that literacy must be seen as a social practice, not an individual act. In other words, interacting with texts is not a single ability that people have or do not have. Instead, people constantly interact with different texts in different socially and culturally appropriate ways. Street believed these varieties to be so diverse that he suggested it made more sense to talk about literacies rather than literacy.

By the early 1990s, these insights led to the “New Literacy Studies,” which viewed literacy not as cognitive skills or personal abilities, but as a set of shared social practices. This body of work has expanded enormously over the last 20 years and now includes many hundreds of publications including notable work by James Paul Gee (1990) and David Barton (1995). It has also profoundly affected school curricula and research agendas in literacy and language education in many parts of the world. The key principle of the New Literacy Studies is recognition of the

diversity of the ways in which people interact with spoken and written language, underlining the futility of considering literacy as a single continuum with the 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 reflects performance of a contextualised ability, and says little about underlying skill or comprehension.

It would, of course, be unfair to criticise a survey instrument designed around 20 years ago for failing to take into account more recent theoretical developments. The main point of the description here is to underline the incompatibility of the instrument with social practices models of literacy, and the need for IALS designers to engage with changing views of the central construct.

The final area of concern is maintaining the validity of the questions in the instrument over time. Each version of the survey has included some questions in common with earlier versions and some new ones. The IALS questions are designed to be contextualised in the lives of adults so that they are less test-like and more authentic. But the lives of adults are very different in different places at different times. Mary Hamilton and David Barton (2000) discuss this at some length. They point out that an exercise involving a bus timetable cannot be universally applied and cannot be interpreted as though it could – European timetables are in 24-h format while US timetables use 12-h format with bolding to represent p.m. Similarly, items become less authentic with the passage of time, as names of software, types of media, presentation of numerical data and other conventions develop over the years.

The IALS tests’ contextualisation was driven by the best possible motives, including assessment theory and a desire to reflect adults’ lives. Unfortunately, the validity of the instrument would have been stronger with a completely decontextualised test. With a question like “What is five times six?”, it would have a great deal of stability. As soon as the same calculation gets put into a specific context, it is vulnerable to changes in that context. So a question such as “Mary has to buy typewriter ribbons for six secretaries. Each uses one ribbon per day, and has half-day Saturdays and Wednesdays. How many does Mary have to buy for a week?” (this is an exaggerated example) is chronologically unstable. The mention of typewriters and secretaries would be unintentional distractors, changing the validity of the question. In addition, respondents would need to know what “half-day” meant in this context, which many contemporary readers would not (that expression is recalled from the 1960s).

The validity problem is a concern with a single survey, but when extended across several surveys it becomes a matter of data reliability. Chronological instability means that the test item is not producing a reliable measure of a specific ability at two different times. The team running the IALS surveys is well aware of this, of course, and has statistical methods for ensuring a degree of compatibility between data sets. What users of the data need to know is that the chronological stability of results cannot be assumed, and that the three IALS surveys do not represent an unproblematically consistent series of measures of the same things. The implication of this is that IALS surveys can only very cautiously be used to measure change in skills within a population. They are not designed for this purpose, and the demands

of rigour within each individual survey actually run against their use as measures of change over time. For example, the weighting of results according to population age will change over time as the proportion of the population in each age band between 16 and 65 rises or declines over time.

If the arguments of this section are accepted, the IALS surveys can be seen as limited in their scope regarding the forms of literacy measured, as reflecting a model of literacy somewhat different from many contemporary models, and as problematic in terms of demonstrating change over time. Accepting these points certainly narrows the range of uses of the data, and raises questions regarding what they can indeed tell us.

### **IALS data: what are they good for?**

The IALS surveys are a remarkable development, very carefully designed to provide robust information on literacy. Throughout the history of these tests there has been tension between those who wished to extend the meaning of the results in all sorts of directions and those (often from within the test team) who have tried to be more cautious. After considerable in-depth study of the issues, it seems reasonable to identify a few aspects of the surveys that should be accepted as important contributions as well as those that must be treated more carefully.

The limitations of the surveys include the model of literacy driving them, which is restricted quite significantly to the ability to retrieve certain types of information from certain types of text and then write a reply in short form or make appropriate indications on a diagram (e.g. circling a number). The claim that this is a robust indicator of an individual person's complex multi-layered set of literacy practices has never been fully discussed and must be treated with caution. The surveys really provide data on one type of activity, namely a particular kind of text consumption in a developed society. The fact that higher IALS literacy levels are associated with, for example, higher income, does suggest that the measure is capturing something of possible economic value, but it is a great leap to assume that that is literacy practices.

The country rankings contained within reports of IALS surveys have always been challenged, and with good reason. There is enough concern about them to set them aside, and it is actually quite unclear how they could be used responsibly anyway. Similarly, the use of IALS surveys as a time series needs great care – their deep compatibility over time remains to be proven. Finally, the claim that level 3 is the functional literacy level for contemporary society is somewhat problematic and should not be reproduced until there is evidence to support it.

It is easy to illustrate the kind of tangles that accepting these aspects of IALS without challenge can lead to. Some jurisdictions have set goals for themselves in terms of IALS levels, such as “By 2020, 70 % of Albertans will have a minimum of level 3 on international adult literacy measures” (Alberta Advanced Education and Technology 2009, p. 9). Here there are implications of comparison over time, acceptance of level 3 (without saying on which scale), and assumptions that the IALS scores provide information on the types of abilities that will be useful to the

economy of the Canadian province of Alberta. Aims expressed in these terms have great political value due to the impression of clarity and action they convey, but are often not helpful at all to those in the field.

Overall, responsible use of the IALS results requires moving away from the expectation that the scales can necessarily be understood as representing different dimensions of literacy, that results can be linked directly to functional literacy levels, or that comparisons between countries and language groups are robust. The value that remains – and it is considerable – is the possibility of linking people’s demonstrated abilities with texts to a wide array of background factors. IALS data demonstrate that age, gender, education, occupation and income are related to individuals’ concrete capabilities, and that they can provide insights into the nature of these structural relationships. They can also tell us which of these matter more, or less, than anticipated (Kalton et al. 1998). This has enormous potential for measuring the effectiveness of educational systems and their contributions to equity. In this use of the results it is less important to know exactly what the surveys measure, because the scores are treated the same for all respondents, so they represent a consistent assessment of something that we know is valuable.

More than this, it should be possible to compare across countries, and to say that gender matters more in one system and less in another. Or that age effects are more significant in an industrialised country and less so in one where agriculture predominates. This provides a unique opportunity for deeply informed critique of educational systems and for the development of adult literacy education strategies grounded not in a claim of economic value but in evidence of the needs for social equity. This application of the IALS data is the most robust – and perhaps the most valuable – use it would be possible to make of the information that has been so painstakingly gathered.

## Conclusion

The aim of this discussion has been to illustrate some of the complexities of the IALS surveys, but not to dismiss them. The purpose was to interrogate them rigorously and point out the misuses and overclaims associated with the surveys, while at the same time identifying some ways in which they can contribute to our understandings of literacy.

The IALS surveys are unique in adult literacy, and indeed in adult learning on a broader scale. They have enormous unexplored potential to provide information on dimensions of education to which they have not yet been applied. They contain some of the best and most detailed evidence available on how the social circumstances within which people live affect one set of their literacy abilities, namely a particular kind of text consumption in a developed society. Yet their value is very significantly undermined by claims that they are an effective way of measuring the distribution of human capital across a society, and more than this, that they provide a means of comparing the population skills of different countries. These claims ignore the limited range of data they take into account, the limits of the IRT model used to convert the raw data into scores, the limited power of the way



levels are derived, and the limited comparability across nations. Applying the survey findings in ways that are not justified does not just lead to ill-informed policymaking, it also obscures the true potential of the surveys.

Given the range and depth of the data, there is the potential to build well-evidenced and insightful analyses of equity in education upon them, and of applying them to the design and delivery of literacy education for adults and children. For example, the different levels of measured skills seen across age cohorts allow for analysis of different evolutions of school structure and for the “erosion” of skills to be better understood. Data on poverty, location, gender, disability, and employment status could be analysed to provide invaluable insights into the details of the way these factors affect people’s lives in different societies. The reason this work is not done is the chimera of comparable human capital data, something that these surveys, due to the assumptions buried within them, will never provide. It would be tragic indeed if this potential remained unrealised, continuing to be obscured by unjustified claims of knowledge about functional literacy, cognitive skills and the nature of reading and writing.

## References

- Alberta Advanced Education and Technology (2009). *Living literacy: A literacy framework for Alberta’s next generation economy*. Edmonton, AB: Government of Alberta.
- Barton, D. (1995). *Literacy: An introduction to the ecology of written language*. Hoboken, NJ: Wiley-Blackwell.
- Becker, G. S. (1975). *Human capital: A theoretical and empirical analysis, with special reference to education*. New York: National Bureau of Economic Research.
- Blum, A., Goldstein, H., & Guérin-Pace, F. (2001). International adult literacy survey (IALS): An analysis of international comparisons of adult literacy. *Assessment in Education: Principles, Policy and Practice*, 8(2), 225–246.
- Carey, S. (Ed.). (2000). *Measuring adult literacy: the international adult literacy survey in the European context*. London: Office for National Statistics.
- Gee, J. P. (1990). *Social linguistics and literacies: Ideology in discourses. Critical perspectives on literacy and education*. London: Falmer Press.
- Hamilton, M., & Barton, D. (2000). The international adult literacy survey (IALS): What does it really measure? *International Review of Education*, 46(5), 377–389.
- Kalton, G., Lyberg, L., & Rempp, J.-M. (1998). Review of methodology. Appendix A. In T. S. Murray, I. S. Kirsch, & L. B. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey*. Washington DC: Office of Educational Research and Improvement.
- Kirsch, I. S., & Guthrie, J. T. (1980). Construct validity of functional reading tests. *Journal of Educational Measurement*, 17(2), 81–93.
- Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: Profiles of America’s young adults. National assessment of educational progress report 16-PL-02*. Princeton, NJ: Educational Testing Service.
- 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., et al. (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.
- Kolstad, A. (1996). *The response probability convention embedded in reporting prose literacy levels from the 1992 Adult Literacy Survey*. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

- OECD (Organization for Economic Co-operation and Development) (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey*. Paris: OECD, Statistics Canada.
- OECD (Organization for Economic Co-operation and Development) (2009). *Adult literacy*. Retrieved June 22, 2009 from [http://www.oecd.org/document/2/0,3343,en\\_2649\\_39263294\\_2670850\\_1\\_1\\_1,00.html](http://www.oecd.org/document/2/0,3343,en_2649_39263294_2670850_1_1_1,00.html).
- Rock, D. (1998). Validity generalization of the assessment across countries. In T. S. Murray, I. S. Kirsch, & L. B. Jenkins (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.
- Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Statistics Canada (2008). *IALS: About the survey*. Retrieved December 1, 2012 from <http://www.statcan.gc.ca/pub/89-588-x/4152886-eng.htm>.
- St. Clair, R. (2010). *Why literacy matters*. Leicester, UK: NIACE.
- St. Clair, R., & Belzer, A. (2007). National accountability systems. In P. Campbell (Ed.), *Accountability in adult basic education*. Edmonton: Grass Roots 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(2), 19–36.
- Street, B. V. (1984). *Literacy in theory and practice*. Cambridge: University of Cambridge.

## The author

**Ralf St. Clair** is Professor and Department Chair at the Department of Integrated Studies in Education at McGill University, Montreal. He has worked in universities and adult education in Canada, the US and the UK. St. Clair has also been involved in projects across Europe, Korea and China, and has wide experience of staff development and training. He has published numerous articles, two jointly edited books, one single authored book and has edited journals. He has been interested in the representation of literacies for some time, and has published and researched extensively in the area.