

## THE INTERNATIONAL ADULT LITERACY SURVEY: HOW WELL DOES IT REPRESENT THE LITERACY ABILITIES OF ADULTS?

Thomas G. Sticht

International Consultant in Adult Education

### Abstract

*In the mid-1990s nations in the Organization for Economic Co-operation and Development (OECD) conducted the first International Adult Literacy Survey (IALS). The IALS used two different methods for assessing adult literacy. One method used performance scales to measure prose, document, and quantitative literacy. The second method measured perceived abilities by having adults rate the extent to which their literacy and numeracy skills met their work and daily life requirements for these skills. This paper reviews evidence that challenges the validity of the IALS standardized performance scales, including the construct validity of the measurement scales (the question of just what it is that the IALS scales measure), the standards validity (the question of how good is good enough to be considered competent at whatever the scales measure), and the use validity (the extent to which the findings are useful for various purposes and do not produce social harm). The author concludes that in future assessments more attention should be given to the use of self-perceptions of skills so those who believe they are in need of additional literacy development can be identified and provided with information about educational opportunities.*

### Résumé

*Au milieu des années '90, les pays membres de l'Organisation de coopération et de développement économique ont mené, selon deux méthodes distinctes, une première enquête internationale sur le niveau d'alphabétisation des adultes. L'une des méthodes choisies a utilisé des échelles de performance pour évaluer de manière quantitative la prose, l'argumentation et l'alphabétisation. La seconde méthode a mesuré les habiletés perçues en demandant aux adultes eux-mêmes d'évaluer à quel point, chaque jour, ils étaient confrontés à leur analphabétisme ainsi qu'à leurs difficultés avec les nombres. Cet article remet en question la validité de ces échelles de performance standardisées pour les adultes, de même que la façon dont ces échelles ont été bâties, c'est-à-dire ; leur objet de mesure et la justesse des critères utilisés. En d'autres mots, à quelle compétence correspond un niveau donné sur cette échelle et quelle utilisation sera faite de ces mesures ultérieurement et enfin, si ces découvertes seront-elles utiles à d'autres usages et n'entraîneront-elles pas de conséquences négatives ? On y conclut que lors de prochaines évaluations, il faudra faire plus attention*

*à l'utilisation des habiletés perçues par les adultes eux-mêmes de sorte que ceux qui pensent avoir besoin de plus de formation puissent être identifiés et qu'on puisse leur procurer de l'information sur les possibilités de s'instruire qui leur sont offertes.*

Early in the 1990s, the Organization for Economic Co-operation and Development (OECD) expressed concerns about the influence of adult literacy on economic competitiveness among member nations. It noted, "one area that is receiving growing attention from educational policymakers and analysts in a number of OECD countries is the direct measurement of literacy levels in the labor force of industrialized countries" (Benton & Noyelle, 1992, p. 11). Several years later, two reports (OECD, 1995, 1997) presented what was to become the first International Adult Literacy Survey (IALS), which eventually involved 22 nations (see Tuijnman, 2000). Similarly, various efforts to define literacy (see Venezky, Wagner, & Ciliberti, 1990) and 75 years of literacy test development (see Sticht & Armstrong, 1994) confirm that there are many ways to approach adult literacy assessment. The IALS researchers used two approaches to assessing adult literacy abilities—standardized performance tests and self-assessments. However, discrepancies arise in comparing the results of the two methods. In this article, analysis of issues raised by other researchers for the IALS reveal questions about the validity and use of the results to construe accurately the nature of literacy and the different ways of constructing a representation of the distribution of literacy in a population.

### **The IALS Survey Methodology and Questions of Its Validity**

One of the methodological innovations of the IALS was the adaptation of a measurement system developed in the United States for use in other nations (OECD, 1995, 1997). The measurement system consisted of three literacy scales: prose, document, and quantitative literacy. The measurement range of scores for each scale was 0 to 500, though in practice scores fell primarily in the 180–390 range of the three scales. For each scale, five levels of literacy were defined, increasing from the lowest level, Level 1 (scores from 0 to 225), through Level 2 (226–275), Level 3 (276–325), Level 4 (326–375), and Level 5 (376–500), the highest level of literacy. The three scales and five levels were formed on the basis of door-to-door surveys in which, among other things, adults in the age range 16 through 65 were asked to perform a number of "real world" tasks involving printed materials and oral instructions.

In addition to the performance task scales used to assess literacy skills, the IALS also created another set of scales which asked adults to provide self-assessments of how well their reading, writing, and numeracy skills met the demands for such skills in their daily lives and at work. The measurement scale for each of these literacy and numeracy skills consisted of five categories: *no opinion*, *poor*, *moderate*, *good*, and *excellent*.

The use of these two different types of measurement methods (performance tasks and self-assessments) resulted in one of the more intriguing findings from the IALS. The number of adults thought to be at risk for various factors—such as low employment, dependency upon welfare, poor health care, lack of civic participation, and so forth—due to low literacy in each nation was much higher when the performance scales were used than when the self-assessment scale was used. For instance, in Canada, on the document scale, 18.2% of adults were assigned to Level 1, the lowest level of literacy based on their performance task results (OECD, 1995, p. 57). This suggests that some 3.3 million of Canada's 18.5 million adults aged 16 through 65 are at risk because of low literacy.

However, on the self-assessment scale of how well they read in daily life, only around 5 percent of Canadian adults (fewer than 1 million) rated their skills in reading for daily life or at work "poor." Of the 3.3 million adults in Level 1—the lowest level of literacy on the document scale—21.9% thought they had excellent reading skills, 26.5% thought they had good reading skills, and 23.9% thought they had moderate reading skills. Fewer than a fourth (22%) of the adults assigned to the lowest literacy level thought their reading skills were poor, and 5.7% had no opinion (OECD, 1995, p. 192). Similar discrepant findings were found for other nations between the IALS performance tests and the self-assessed reading abilities for the two other literacy scales and for self-assessments of writing and numeracy skills.

In the IALS, the performance scales and the self-assessments represent two fundamentally different approaches to assessing adult literacy abilities. In the performance assessments, literacy is construed as a cognitive ability (latent trait) that makes possible the use of printed materials in various contexts. It is considered that some people have more of this capacity than others, although how much people have or lack may not be consciously apparent to them. Nonetheless, it is assumed that these differences in the amount of capacity can be inferred using people's performance on various real-world tasks that incorporate the latent trait that is theorized to make possible each person's performance.

In the self-assessment approach to assessing literacy, literacy is considered as an ability or set of abilities (as in reading, writing, and numeracy in the IALS) that adults are consciously aware of and can perceive well enough to estimate how well their literacy skills permit them to negotiate the literacy demands of different sets of activities at work or in their daily life. This requires that adults are aware both of the demands for literacy in the different contexts that they encounter and of how well their literacy abilities permit them to meet these demands on a recurrent basis.

Clearly, these two different approaches to assessing adult literacy are based on different implicit theories about literacy and different procedures for measuring literacy. It is also evident from the discrepancies in data that these approaches produce different estimates of how many adults are at risk because of literacy in the various nations that participated in the IALS. These

findings raise serious questions about the validity of the different assessments: Is each assessment equally valid as a means for representing the literacy abilities of the adult population? If so, then how should the different results of each method be used? In the following analysis, the task performance method as used in the IALS is examined first and issues are discussed about its validity in representing adult literacy. Following that, the validity of the self-assessment method for assessing the literacy abilities of adults is examined.

### **Validity Issues With the IALS Performance Scales**

Assessors of adult literacy over the decades have constructed various representations of adult literacy, and these different representations of adult literacy often produce discrepant findings, such as those noted above. Thus, Benton and Noyelle (1992) conclude, "the direct measurement of literacy levels in the labor force of industrialized countries" (p. 11) is not possible, because literacy per se does not exist, in the labor force or any where else for that matter, as something to be directly measured. Instead, different representations of literacy may be created based on different constructs and theories of what literacy is and why it should be represented in one way rather than another. Major challenges to the validity of the IALS standardized performance test scales include (a) the construct validity of the measurement scales, that is, the question of just what it is that the IALS scales measure, (b) the standards validity, that is, the question of how good is good enough to be considered competent at whatever the scales measure, and (c) the use validity (Messick, 1989), that is, the extent to which the findings are useful for various purposes and do not produce social harm.

#### ***Construct Validity: What Do the IALS Scales Measure?***

In the United States over the last 80 years a variety of different representations of adult literacy have been socially constructed by psychometricians, statisticians, and survey experts in consultation with various stakeholder groups including adult educators, adult students, literacy researchers, policymakers, and others with a declared interest in adult literacy (Sticht & Armstrong, 1994). For example, the U.S. National Assessment of Educational Progress of 1970–1971 assessed young adults' (26–35 years of age) literacy using the same academic tasks that were used to assess the growth of literacy in school children aged 9, 13, and 17 years. This involved assessments in skills that reading teachers have traditionally taught, including knowledge of word meanings (vocabulary), using visual aids, following written directions, using reference materials, locating significant facts, getting the main idea from materials, drawing inferences, and critical reading (Sticht & Armstrong, p. 128). In contrast to such academic tasks, functional, real-world tasks were used in the Survival Literacy survey by Louis Harris Associates in 1970, the Adult Functional Reading Study of 1973, the Adult Performance Level Study of 1975, the

Young Adult Literacy Survey (YALS) of 1985 and the National Adult Literacy Survey (NALS) of 1993.

The last two assessments, the YALS and NALS, provided the methodology for the development of the performance tasks and measurement scales used in the IALS. The NALS methodology was developed after the U.S. Congress passed the Adult Education Amendments of 1988 that required the U.S. Department of Education to submit a report to Congress on the definition of literacy and then to report on the nature and extent of literacy among adults in the nation (Campbell, Kirsch, & Kolstad, 1992, p. 2). With the aid of a national advisory board, the National Center for Education Statistics (NCES) agreed to define 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, Jungeblut, Jenkins, & Kolstad, 1993, pp. 2-3).

The advisory panel for the NALS also agreed on the definitions of three literacy scales that were developed to represent the literacy skills of adults. These included prose literacy, the knowledge and skills needed to understand and use information from texts; document literacy, the knowledge and skills required to locate and use information contained in various formats; and quantitative literacy, the knowledge and skills required to apply arithmetic operations embedded in printed materials (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993, p. 3). These same definitions were used in the IALS (OECD, 1995, p. 14).

In both the NALS and the IALS, there was an important conceptual shift in going from the general definition of literacy (using printed and written information to function in society, etc.) to the definition of the three literacy scales that were actually used to characterize adult literacy in the various countries. None of the definitions of the three scales started with *using printed and written information* as the defining aspect of literacy. Rather, all three definitions referred to the cognitive constructs of *the knowledge and skills* or *the ability* required to perform tasks using prose, document, and quantitative materials. This makes the important shift in conceptual understanding that, unlike what the general definition of literacy states, literacy in the IALS was not considered as the *use* of printed materials, but rather the *knowledge and skills* or *ability* that make possible the use of printed materials for various purposes.

This important shift in the conceptual framework for the IALS was not addressed by Kirsch, Jungeblut, and Mosenthal (1998) in their tracing the history of the theoretical framework used to construct the IALS. Instead of focusing on the knowledge and skills underlying performance on the three measurement scales, they note that the framework used to develop assessment tasks was based on the types of materials involved in the assessments, the uses that respondents were asked to make of the materials (i.e., the types of tasks performed or questions to be answered) and what they

refer to as "three important and distinct areas: prose, document, and quantitative literacy" (p. 106).

In a number of analyses, Kirsch and Mosenthal (1990) and Mosenthal and Kirsch (1991) found that the difficulty of tasks (e.g., percentage getting various items correct) was a function of the information processing complexity of the task (including things like the numbers of items from the questions or task instructions that must be held in mind while searching through materials to locate answers) the amount of materials to be searched or read, the plausibility of distracting information as being the target information in materials, the abstract versus concrete nature of the language in the materials, or the types of and amounts of calculations that had to be done in the quantitative literacy tasks. The first report of the IALS findings (OECD, 1995) states, "The IALS builds on the seminal work of Kirsch & Mosenthal (1990) with respect to adult reading. The IALS exploits their theoretical framework, which explains the factors that underlie difficulty in adult reading" (p. 25). Although it is true that there are empirical data to support the usefulness of the various material and task variables that the researchers identified through their information processing analyses to predict the difficulty of prose, document, and quantitative tasks, when the latter were scaled for difficulty using the IALS methodology, there is no explicit discussion of a cognitive theory of the competence (knowledge and skills) that adults must possess to be able to perform the many tasks.

Kolstad et al. (1998) present data that raises questions on the construct validity of the Kirsch and Mosenthal (1990) theory underlying the IALS. They note that Kirsch, Jungeblut, and Mosenthal (1994) used the IALS scale of difficulty based on respondents' having an 80% probability of getting items correct to validate their analysis of what made items less or more difficult and to determine at what points on the 0 to 500 point measurement scales the scales should be divided into the five levels of difficulty. But, when lower response probabilities were used by Kolstad et al. to test the robustness of the Kirsch and Mosenthal factors for predicting the difficulty of test items, the factors contributing to the difficulty of items changed. At the 80% standard used by Kirsch, Jungeblut, and Mosenthal the plausibility of distracting words or sentences in the stimulus materials was significantly related to the difficulty of the tasks, whereas estimates of difficulty using a readability formula were not significant. But at response rates of 50% or lower, just the opposite was found; readability formula estimates of difficulty were significant but plausibility of distracting material was not (Kolstad, et al., p. 23). This challenges the construct validity of the Kirsch and Mosenthal theory of literacy underlying performance on the IALS, because the factors that determined the difficulty of the IALS items changed not as a function of a change in some specified aspect of literacy but rather as a function of the standard of proficiency that was adopted to be considered proficient in certain tasks. (This point is discussed further in the subsection on standards validity, below.)

Additional evidence exists to question the construct validity of the three IALS scales as distinct scales. In independent analyses, Reder (1998) and Rock (1998) found that the prose, document and quantitative scales correlated around  $+0.90$ , sharing some 80% of their variance, yet there is no theoretical statement about cognition that might account for this large overlap among the three literacy scales. These findings support arguments against the construct validity of the three scales as distinct, as expressed by the theory underlying the IALS (Kirsch, Jungeblut, & Mosenthal, 1998, p. 106).

Much of the commonality among the three scales could result from all three scales' reliance on knowledge and language components (words, syntactical rules) that overlap. The importance of knowledge in literacy assessment was acknowledged by the advisory panel for the NALS in its acceptance of the definitions of the three different literacy scales (prose, document, quantitative) that were developed. In each case, literacy was defined as the knowledge and skills needed to perform the three types of literacy tasks. From these definitions, it is clear that the advisory panel for the NALS understood that the use of printed and written information to accomplish tasks requires certain underlying knowledge and skills to make such use possible. Yet in the IALS there was neither measurement of what people knew about the various subject matters sampled in the various tasks—for instance the vocabulary or conceptual knowledge used in the tasks—nor whether they could use their literacy skills to increase their knowledge of various subject matters.

The literacy survey data for the IALS show a large decline of performance for older adults (OECD, 1997) even when the data are adjusted for differences in the years of education of younger and older adults. Because working memory becomes increasingly less efficient with advanced age (see Bernstein, Roy, Srull, & Wickens, 1988; Meyer, Marsiske, & Willis, 1993), these findings strongly suggest that the IALS tasks derived a great deal of their difficulty from the load they placed upon working memory, a factor that could threaten their validity as literacy measures for the elderly.

Because the IALS does not measure knowledge of a vocabulary or cultural nature, both of which have been demonstrated to increase with age (Hofstetter, Sticht, & Hofstetter, 1999; Sticht, Hofstetter, & Hofstetter, 1996) and instead emphasizes "search and locate" types of tasks that introduce unknown and possibly irrelevant test variance due to the overloading of working memory, the construct validity of the performance assessments is questionable for older adults. The IALS may produce serious underestimations of the breadth of materials that older adults can read and comprehend (using their more extensive, and in some cases specialized knowledge base) and the tasks they can perform (given sufficient time to study materials) without the pressure for efficiency typical of test-taking situations. Such test situations are of questionable real-world validity in the lives of most adults over the age of 25 who are not in school.

This sort of ambiguity about what is being assessed when so-called real-world tasks like those of the IALS are used in adult literacy assessments has long been recognized. This is clearly illustrated in a manual for item writing produced by the Comprehensive Adult Student Assessment System (CASAS, 1983). The CASAS is an adult literacy assessment system widely used in the United States and in some other industrialized nations that uses functional, real-world literacy tasks much like those of the IALS. The manual notes that the use of complex, real world, or functional tasks as items "generally tests the use of two or more skills. Therefore, this context is not appropriate in itself for diagnosing weaknesses in specific skills since it is difficult to determine which skill was performed incorrectly" (p. 1). A decade later, Venezky (1992) reiterated this point about the ambiguity of assessments using real-world tasks; he notes that functional literacy tests (like those of the IALS) may lack construct validity because they are not derived from theoretical models of the knowledge and skills involved in literacy. Instead, such tests use tasks that engage complex information processing activities with unknown mixtures of various knowledge and processes. For this reason, he concludes, it is not clear what they assess nor what their instructional implications are (p. 4).

In the absence of a clearly specified theory of literacy as a psychological construct, it is not possible to know how to develop assessments that measure the component knowledge and skills that make up the ability (or abilities) that constitute literacy. Without knowing what specific knowledge or skills are being assessed in real-world tasks, it is not clear to what extent test performance reflects literacy ability or some other abilities (such as problem solving, reasoning, language comprehension, vocabulary knowledge, management of test-taking anxiety, interpersonal skills, motivation, competitiveness, or some complex, interactive combination of all these or whatever). That is why in the YALS, the earlier version of the NALS and IALS, the test items were referred to as "complex information processing" tasks (Kirsch & Jungeblut, 1986).

### ***Standards Validity: How Good Is Good Enough?***

As noted above, in the IALS assessment adults' perceived literacy abilities exceeded their abilities as estimated by performance test results. Why might this have happened? One answer can be found in the way the IALS defined proficiency at a given level. The IALS developers set a criterion of having an 80% probability of getting the average item at a given literacy level correct to be considered proficient at that level of skill. However, Andrew Kolstad, the leader of the National Adult Literacy Survey (NALS) project at the National Center for Education Statistics in the United States, which produced the technical approach to measurement used by the IALS, later argued that the 80% response probability level was arbitrary (Kolstad, 1996). He notes that the .65 response probability standard is used by the National Assessment for Educational Progress (NAEP) for children in the public school system in the United States. Using NALS data, he then

recalculated the percentage of adults who would be in literacy Level 1 if a standard of 65% probability of a correct response was used for adults. In this case the percentage of adults assigned to Level 1 on the prose scale fell from around 20% to 13%.

Kolstad (1996) also determined the consequences of using the 50 percent probability criterion used by the Comprehensive Adult Student Assessment System (CASAS), which is an adult literacy assessment used in the United States and other nations and is included in the United States federal government's dissemination network for effective educational innovations. Using the 50 percent response probability standard with NALS data, only 9% of U.S. adults fell in Level 1. This suggests that if Kolstad's findings using a 50% probability criterion were applied to the IALS data in each country this would result in about a 50% reduction of adults considered most at risk for literacy.

In later analyses, Kolstad et al. (1998) conducted analyses of the NALS data and demonstrated empirically that the use of the 50% response probability standard produces the least errors in determining whether adults can or cannot perform literacy tasks across the full range of tasks included in the NALS or IALS. On the question of how good is good enough, they give an answer based on the statistical theory used to construct the IALS performance assessments. They conclude that the 50% standard should be used in these adult literacy assessments because it would "make the best use of the test information" whereas the 80% response probability "provided a poor approximation of the latent trait [i.e., presumably literacy] required to perform successfully on the test items" (p. 51).

For the nations that have used the IALS it is important to realize that the decision to set the standard of performance at the 80% response probability level to categorize adults into one of the five levels of literacy on the IALS was an essentially arbitrary decision about what constitutes competence (or mastery) of literacy at different levels. Adults with a score of 200 on the prose literacy scale were assigned to the lowest level of literacy, Level 1, because they had an 80% probability of being able to perform the average task at prose Level 1. However, as the IALS developers note, the same adults would be expected to be able to perform 40% of the average Level 2 tasks, 18% of the average Level 3 tasks and even 6% to 7% of the average tasks at the most difficult Levels of 4 and 5 (OECD, 1997, p. 132, Table A.1). Thus, in the IALS, although the use of the 80% response probability standard greatly reduced the probability of making false positive errors (that is, saying people could perform tasks who actually could not), it greatly increased the probability of making false negative mistakes (that is, saying that people could not perform tasks when in fact they could).

Similarly, using an arbitrary response probability of .80 for the document and quantitative scales resulted in millions of adults being assigned to the lowest levels of literacy when the IALS data actually estimated that they could be expected to perform significant percentages of

tasks well above their assigned level. Exactly how this ability above the adult's assigned level of competence was supposed to be construed as a part of their competence was not addressed, other than to say that they could not "consistently" perform tasks at a higher level as they could those of their assigned level. But it seems likely that this "inconsistent ability" may explain, at least in part, why most of the adults in Level 1 of the IALS literacy scales thought their literacy skills were just fine for meeting the requirements of their work or daily lives. They actually were capable of performing quite a few items at a higher level of literacy.

### ***Use Validity: Consequences in Using the IALS Performance Task Information***

Messick (1989) questions the uses of test scores, including detrimental uses, in the facet of validity that he calls *action* or *use* inferences. He states, "To validate an *action inference* [italics added] requires validation not only of score meaning but also of value implications and action outcomes, especially of the relevance and utility of the test scores for particular applied purposes and of the social consequences of using the scores for applied decision making" (p. 5).

Unfortunately, there were negative social consequences when the NALS report came out in 1993 in the United States and revealed that some 47% to 48% of adults scored in the two lowest literacy levels. This led to newspaper articles around the country with headlines such as that appearing in The San Diego Union-Tribune of Thursday, September 9, 1993: "*Illiteracy hurts half of adults.*" But this was so unbelievable that even political cartoonists jumped on the bandwagon and produced numerous cartoons making fun of the nation's one-half who were allegedly functional illiterates (e.g., *Newsweek*, September 20, 1993). Reinforcing the seeming absurdity of the NALS data, there were no dramatic new programs or increases in funding for adult literacy education announced by the federal government that had funded, conducted, and announced the results of the study. Seven years later, the last report of analyses of the IALS data funded by the United States was unveiled at a meeting of adult educators in Washington, DC (Tuijnman, 2000). But this time no coverage of the reports findings by major newspapers around the nation were found.

In the United States, probably the most important question that NALS researchers were asked to report on was, "Are the literacy skills of America's adults adequate ... to ensure individual opportunities for all adults, to increase worker productivity, or to strengthen America's competitiveness around the world?" (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993, p. xviii). The NALS report answered the question as, "Because it is impossible to say precisely what literacy skills are essential for individuals to succeed in this or any other society, the results of the National Adult Literacy Survey provide no firm answers to such questions" (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993, p. xviii).

Similarly, the IALS reports also provide no firm answers to these same critical questions for the various nations involved in the survey, though it is generally implied or straightforwardly asserted that "Level 3 is regarded by many experts as the minimum level of competence needed to cope adequately with the complex demands of everyday life and work; after all, knowledge societies will dominate the twenty-first century" (OECD, 1997, p. 3). But none of the IALS reports actually identify who the "many experts" are and what methodologies they have used to estimate whether or not Level 3 literacy skills are needed to cope with "the complex demands of everyday life." Some researchers using the IALS data have suggested that millions of adults possess literacy skills that *exceed* their work requirements. For example, Krahn and Lowe (1999) devised a method of comparing the literacy skills of Canadians on the IALS with their reported use of literacy on their jobs. They report that 20% of Canadians are employed in jobs that do not take advantage of their literacy skills. In fact, the percentage of overskilled exceeded the percentage of estimated underskilled adults in the workplaces of Canada. The IALS reports have not addressed this issue of the potential of a literacy "surplus" in the knowledge societies of the 21st century. But this may have something to do with why many businesses and industries have not gotten involved in workplace literacy provision (Blunkett, 2001).

The construct validity problem with the IALS leads to a lack of understanding of what it is that is being measured; however, from the perspective of use validity, this may not be so important for those who perform well, and who can therefore be assumed to possess whatever knowledge and skill is called for in performing the "complex information processing" (literacy?) tasks. But the problem becomes critical when the focus of concern is on understanding why it is that those who do not perform well do not perform well and what should be done about it. What kinds of services should be provided to help them improve their ability to perform these kinds of tasks? Is literacy education called for, or would training in memory enhancing techniques, logical reasoning, and critical thinking be the thing to do? How much of the problem is a *literacy* problem in contrast to a general education problem?

The National Center for Education Statistics (1998) report a meeting of experts in Washington, DC that was convened to discuss the problem of the arbitrary nature of the .80 response probability for defining competence at a given literacy level on the NALS and IALS. In the portion of the document summarizing the outcomes of the meeting, they cite Andrew Kolstad, project director for the NALS, who emphasized:

The most frequent policy use of the findings is on reporting the high percentages of adults who do not reach the higher levels. Inferences about the low end of the scale, about what people can't do, should put more weight on false negatives (being sure that those we say can't do certain tasks really fail to perform the tasks). The priorities of the adult

literacy program really should dictate a value below .50 for the RP [response probability] convention, in order to ensure that inferences about substandard performance are on solid ground." (p. 13)

The report contained comments by other experts acknowledging the political aspects of changing the arbitrary standard for proficiency in literacy from the 80% to the 50% response probability level, even if the latter is the statistically appropriate standard. According to two members of the group, "The problem is largely one of 'public relations.' It is easy to agree with the statistical arguments, but how do we explain to the public that RP50 is in some sense sufficient" (p. 10).

In short, what is seen in the discussion of the experts is that the test developers understand the arbitrariness of the IALS scaling, yet they disagree about what, exactly, should be done about it. They are concerned that the media or public will not understand why the .50 response probability standard might be more appropriate than the one used in the surveys. As one expert put it, "A great deal of effort needs to be expended on trying to make a statement the public can understand." And another said, "We need to think how we can protect NCES products from misinterpretations by media and other users" (National Center for Education Statistics, 1998, p. 10). These comments raise serious doubts about the use validity of the IALS performance task assessments for usefully representing the literacy abilities of adults and their needs for literacy education services to policymakers, adult educators, the media, and the public at large.

### **Validity Issues With the IALS Self-Assessment Scales**

Different representations of literacy may be created based on different ideas of what literacy is and why it should be represented in one way rather than another. In the latent-trait approach to literacy assessment used by the IALS performance tasks, people's abilities were determined by first using sample data to compute difficulty levels for various tasks and then computing people's ability levels from their performance of the tasks that had been scaled for difficulty. In this case, estimates of a given adult's literacy abilities were highly dependent on the literacy abilities of the other adults in the samples used to construct the scales for prose, document, and quantitative literacy. This means that in the performance assessment methodology, an adult's literacy ability is defined as a *social construct*—that is, one's ability is estimated not in some form of absolute terms, but rather in terms of how well one performs on tasks that have been scaled for difficulty by having samples of adults try to perform the various tasks.

In the self-assessment approach used by the IALS, people were asked to assess their ability levels based on their judgments of how well their literacy skills met the demands of their work or other daily needs for such skills. So in this case literacy was not considered as an ability relative to social group norms but rather as an individual ability that met or did not meet the perceived daily requirements for such skills in two different contexts, at

work or in daily life. Had respondents been asked to rate their reading, writing, and numeracy skills in terms of whether or not they were poor, moderate, good, or excellent readers compared to other readers in the social groups they typically associated with, then the scale of literacy would have been a socially based scale, and different results might have been obtained. Recent work in the United Kingdom (Basic Skills Agency, 2001) has demonstrated additional types of self-assessments of adult basic skills, all of which must be evaluated with regard to the question of what type or types of validity they exhibit.

### ***Construct Validity: What Do the IALS Self-Assessment Scales Measure?***

The IALS reports do not provide any specific definitions used for reading, writing, and numeracy in the self-assessments of literacy and numeracy skills. Apparently, it was left to each of the respondents to create an understanding of the three terms and to respond with their ratings of their abilities based on their self-constructed understandings. However, Messick (1989) refers to construct validity as the validity of interpretive inferences; he states, "To validate an *interpretative inference* [italics added] is to ascertain the extent to which multiple lines of evidence are consonant with the inference, while establishing that alternative inferences are less well supported. This represents the fundamental principle that both convergent and discriminant evidence are required in test validation" (p. 5).

Using this approach to construct validity, it can be assumed for purposes of analysis that the self-assessment and the performance test methods have equal construct validity. It is then possible to compare the self-assessment and performance testing approaches in terms of how well each approach predicts the outcomes of the other. For instance, using the performance test data as the predictor variable and the self-assessment data as the criterion variable it is possible to determine how well the performance tests predicted the self-assessment results. As indicated earlier using the data for Canada as an example, the document scale performance assessment methodology assigned some 3.3 million adults into literacy Level 1 as poorly literate. However, only 22% of the adults assigned to Level 1 actually considered themselves as poor readers and over 60% to 70% thought they were moderate, good, or excellent at meeting the reading requirements for their daily lives and work. In this case then, the performance task methodology was not very accurate in predicting the self-assessment results. It produced millions of false negatives—that is, people who were declared to be poorly literate whose self-ratings said they were actually moderately to excellently literate.

When the self-assessment data are used as the predictor variable it is found that only some 5% of Canadian adults rated themselves as poor in meeting their daily needs for reading and work. Of these adults, over 78% were found in the lowest level of literacy as defined by the document performance test criterion variable, resulting in a much lower incidence of

false negatives (i.e., adults who said they did not read well who actually performed well on the IALS) than the performance test methodology produced. Even given that many adults may have over-estimated their reading skill levels on the self-assessments, creating numerous false positives (i.e., adults who said they could read well who actually did not perform above Level 1 on the IALS), this analysis lends support to the interpretive inference (construct validity) that the self-assessment scale provides a valid measure of literacy ability.

### *Use Validity of the Self-Assessments*

Interestingly, the IALS researchers appear to reject the validity of the self-assessments as indicators of the adult's literacy abilities. In the first of the IALS reports, in a discussion of the self-assessment findings, the OECD (1995) states in a side bar that "most people do not recognize that they have a literacy problem" (p. 101). Later the report states,

Numerous studies have shown how adults with low literacy skills are able to construct their daily lives so that literacy is not a part of it and therefore, they can legitimately claim that their skills serve them well. These coping mechanisms often lead to individuals being dependent on others to meet their literacy needs. (p. 109)

No references for the "numerous studies" referred to are cited nor were data found in any IALS reports regarding the extent to which adults in literacy Level 1 received help from others in accomplishing literacy tasks in their daily lives. However, in the United States, data from the NALS indicate that only some 14% to 25% of adults in Level 1 reported that they received "a lot of help from family members or friends with everyday prose, document, and quantitative tasks" (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993, p. xv). These figures are similar to the percentages of adults in Level 1 in the United States who rated themselves "poor" in reading on the three performance scales (OECD, 1995, pp. 191-193).

In some OECD nations efforts are underway to provide literacy education to adults. This has led to the use of the IALS data to indicate the scale of need for adult literacy education in these nations. In setting the scale of need, there are considerable differences in the percentages of adults considered to be at risk and hence in need of literacy or basic skills instructional services based on IALS performance test data and the actual numbers of adults who seek out literacy instruction and enroll in programs. In Canada, for instance, Long (2001) reports, based on IALS data, that some 22% of Canadians have serious difficulties with any type of printed material, yet only a small fraction (5% to 10%) of adults eligible for literacy education have ever enrolled in literacy courses. In the United Kingdom the IALS assigned around 23% (over 7,000,000) adults to literacy Level 1 (OECD, 1997, p. 151) while participants in adult literacy programs around that time included fewer than 5% of that number (Blunkett, 2000, p. 4). Similar discrepancies between those declared in need by the IALS performance tests

and those who actually enroll in basic skills programs have been found in the United States (National Literacy Summit 2000, p. iii).

Although self-assessments may include overestimations of their skills by adults, there is nonetheless a closer agreement between the number of adults who rate themselves as "poor" in literacy in the IALS and the number who enroll in programs than between the IALS performance test results and those who enroll in programs. If, as seems likely, adults are more likely to enroll in adult literacy programs if they think they have poor reading skills, then self-assessments may provide greater use validity than performance tests for determining how many adults may be encouraged to seek out literacy provision because of their poor literacy and numeracy skills.

### Conclusions

Returning to the question in the title of this article, how well does the IALS represent the literacy levels of adults, the evidence appears mixed. On one hand, the analysis of the data for the construct validity, standards validity, and use validity of the IALS performance test data raises serious questions about the validity of these performance assessments. On the other hand, the self-assessment data suggest that such assessments may be more useful for identifying adults who actually believe themselves to be at risk because of poor literacy skills.

However, it is entirely possible that neither the performance test measures nor the self-assessment measures accurately represent how well adults use their literacy abilities in day-to-day situations. Hamilton and Barton (2000) and Gomez (2000) offer an extended discussion of the decontextualized conditions under which the IALS was administered and the artificial nature of the IALS performance test items. Their work challenges the ecological validity of the findings for use in cross-national comparisons or for the formation of adult education policy aimed at actually meeting adult learning needs in the contexts in which they live.

The lack of a sound theory of literacy has led to the development of adult literacy assessments over the last three decades using a variety of different materials and tasks, all with questionable construct validity (Sticht & Armstrong, 1994). If a well formulated, validated theory of the knowledge and skills that make up literacy could be developed, it might be possible to more directly and cost-effectively assess that knowledge and those skills and to predict the performance of a wide variety of both academic and real-world literacy tasks that go beyond those used in present assessments.

Although such a comprehensive theory has not been found, there has been some effort by the National Center for Education Statistics in the United States to look more broadly at theories of literacy that might be used to develop methods for assessing adult literacy skills using some form of performance measures. For instance, in another paper (Sticht, 2000) I discuss the theory of reading, as a second signaling system for speech, to be a separate information processing task from that of searching and locating

written information in graphic displays such as forms, lists, and other types of documents. Focusing on these different aspects of literacy might make it possible to discover whether people have problems comprehending the written version of their spoken language, in contrast to problems in searching and locating information in visual displays of varying complexity and unfamiliarity.

At this time, many nations are using the results of the IALS performance tests to argue that large percentages of adults are lacking in basic literacy and numeracy skills and that nations need to commit large sums of money to greatly enlarge the extension of basic skills provision for adults. At the same time however, these nations are acknowledging that most adults whom the IALS says are in need of basic skills education based on the IALS performance tests do not think they have poor skills and are not presenting themselves for service (Long, 2001, Blunkett, 2000).

The IALS assessment methodology did not permit feedback to adults regarding their performance scores, so this information was not available to them to use in making judgments as to whether they should seek basic skills provision. There are many psychological, institutional, and material conditions of deprivation that may constitute serious barriers to participation in adult literacy education (see Beder, 1991; Blunkett, 2000; Long, 2001; Quigley, 1997). This suggests that, in future national assessments, it may be important to pay more attention to determining adults' perceived levels of literacy and numeracy and providing those who perceive themselves as poorly skilled with information that can be useful in assisting them to overcome various barriers to participation and may motivate them to seek out educational provision.

### References

- Basic Skills Agency. (2001). *Getting better basic skills*. London: Basic Skills Agency.
- Beder, H. (1991). *Adult literacy: Issues for policy and practice*. Malabar, FL: Krieger.
- Benton, L., & Noyelle, T. (Eds.). (1992). *Adult illiteracy and economic performance*. Paris: Organisation for Economic Co-operation and Development, Centre for Educational Research and Innovation.
- Bernstein, D., Roy, E., Srull, T., & Wickens, C. (1988). *Psychology*. Boston: Houghton Mifflin.
- Blunkett, D. (2000). *Skills for life: The national strategy for improving adult literacy and numeracy skills*. London: Department for Education and Employment.
- Campbell, A., Kirsch, I. S., & Kolstad, A. (1992, October). *Assessing literacy: The framework for the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics.
- Comprehensive Adult Student Assessment System. (1983). *The CASAS manual for item writing*. San Diego, CA: Author.
- Gomez, S. V. (2000). The collective that didn't quite collect: Reflections on the IALS. *International Review of Education*, 46, 419-431.

- Hamilton, M., & Barton, D. (2000). The International Adult Literacy Survey: What does it really measure? *International Review of Education*, 46, 377-389.
- Hofstetter, R., Sticht, T., & Hofstetter, C. (1999). Knowledge, literacy, and power. *Communication Research*, 26, 58-80.
- Kirsch, I., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.
- Kirsch, I., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993, September). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: US Government Printing Office.
- Kirsch, I., Jungeblut, A., & Mosenthal, P. (1994, March). *Moving toward the measurement of adult literacy*. Unpublished paper presented at the National Center for Education Statistics conference on literacy levels. Washington, DC: National Center for Education Statistics.
- Kirsch, I., Jungeblut, A., & Mosenthal, P. (1998, January). The measurement of adult literacy. In T. Murray, I. Kirsch, & L. Jenkins, *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey* (pp. 105-134). Washington, DC: National Center for Education Statistics.
- Kirsch, I., & Mosenthal, P. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25, 5-30.
- Kolstad, A. (1996, April). *The response probability convention embedded in reporting prose literacy levels from the 1992 National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics.
- Kolstad, A., Cohen, J., Baldi, S., Chan, T., DeFur, E., & Angeles, J. (1998, May). *The response probability convention used in reporting data from IRT assessment scales: Should NCES adopt a standard?* Washington, DC: National Center for Education Statistics.
- Krahn, H., & Lowe, G. (1999, Summer). Literacy in the workplace. *Perspectives*, pp. 38-44.
- Long, E. (2001, January). Patterns of participation in literacy and upgrading programs: Key results from the national study. *Literacy at Work*, 32, 6-7.
- Meyer, B. J. F., Marsiske, M., & Willis, S. L. (1993). Text processing variables predict the readability of everyday documents read by older adults. *Reading Research Quarterly*, 28, 235-248.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5-11.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33-45). Hillsdale, NJ: Erlbaum.
- Mosenthal, P. & Kirsch, I. (1991). Toward an explanatory model of document literacy. *Discourse Processes*, 14, 147-180.
- National Center for Education Statistics. (1998, September). *The response probability (RP) convention used in reporting data from IRT assessment scales: Should NCES adopt a standard? Meeting of experts*. Washington, DC: US Department of Education.
- National Literacy Summit (2000, September). *From the margins to the mainstream: An action agenda for literacy*. Washington, DC: National Institute for Literacy.

- Organization for Economic Co-operation and Development. (1995). *Literacy, economy, and society*. Paris: Author.
- Organization for Economic Co-operation and Development. (1997). *Literacy skills for the knowledge society*. Paris: Author.
- Quigley, B. A. (1997). *Rethinking literacy education: The critical need for practice-based change*. San Francisco: Jossey-Bass.
- Reder, S. (1998). Dimensionality and construct validity of the NALS assessment. In M. C. Smith (Ed.) *Literacy for the twenty-first century* (pp. 37-57). Westport, CT: Praeger.
- Rock, D. (1998, January). Validity generalization of the assessment across countries. In: T. Murray, I. Kirsch, & L. Jenkins, *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey* (pp. 135-142). Washington, DC: National Center for Education Statistics.
- Sticht, T. (1998, September). *Beyond 2000: Future directions for adult education*. Washington, DC: US Office of Education, Division of Adult Education and Literacy.
- Sticht, T. (2000, March). *Using telephone and mail surveys as a supplement or alternative to door-to-door surveys in the assessment of adult literacy*. Working Paper No. 2000-06. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Sticht, T. G., & Armstrong, W. B. (1994, February). *Adult literacy in the United States: A compendium of quantitative data and interpretive comments*. Washington, DC: National Institute for Literacy.
- Sticht, T. G., Hofstetter, C. R., & Hofstetter, C. H. (1996). Assessing adult literacy by telephone. *Journal of Literacy Research*, 28, 525-559.
- Tuijnman, A. (2000, September). *International Adult Literacy Survey: Benchmarking adult literacy in America: An international comparative study*. Washington, DC: US Department of Education.
- Venezky, R. L. (1992, May). *Matching literacy testing with social policy: What are the alternatives?* Philadelphia: National Center on Adult Literacy.
- Venezky, R. L., Wagner, D. A., & Ciliberti, B. S. (Eds.) (1990). *Toward defining literacy*. Newark, NJ: International Reading Association.