

International Adult Literacy Survey (IALS)

Website: <http://nces.ed.gov/Surveys/all/results.asp>

1. OVERVIEW

The 1994 International Adult Literacy Survey (IALS) represented a first attempt to assess the literacy skills of entire adult populations in a framework that provided data comparable across cultures and languages. This collaborative project was designed to inform both education and labor market policy and program development activities in participating countries. The international portion of the study was carried out under the auspices of an International Steering Committee chaired by Canada, with each participating country holding a seat on the committee along with representatives from the Organization for Economic Cooperation and Development (OECD), European communities, and the United Nations Educational, Scientific and Cultural Organization.

In the United States, IALS is the fourth assessment of adult literacy funded by the federal government and conducted by the Educational Testing Service (ETS). The three previous efforts were (1) the 1992 National Adult Literacy Survey (see [NALS chapter](#)); (2) the Department of Labor's (DOL) 1990 Workplace Literacy Survey; and (3) the 1985 Young Adult Literacy Assessment (funded as an adjunct to the National Assessment of Educational Progress—see [NAEP chapter](#)). In order to maximize the comparability of estimates across countries, IALS chose to adopt the National Adult Literacy Survey methodology and scales. Literacy was defined along three dimensions—prose, document, and quantitative. These were designed to capture an ordered set of information-processing skills and strategies that adults use to accomplish a diverse range of literacy tasks encountered in everyday life. The background data collected in IALS provide a context for understanding the ways in which various characteristics are associated with demonstrated literacy skills.

IALS was originally conducted in eight countries (Canada, Germany, Ireland, the Netherlands, Poland, Sweden, French- and German-speaking Switzerland, and the United States). A second phase was subsequently conducted in five additional countries or territories (Australia, Flemish-speaking Belgium, Great Britain, New Zealand, and Northern Ireland), and in a final phase included an additional nine countries. This chapter focuses on the first phase, in which the United States participated.

Purpose

To (1) develop scales that would permit comparisons of the literacy performance of adults (16 and older) with a wide range of abilities; (2) if such an assessment could be created, describe and compare the demonstrated literacy skills of adults in different countries.

1994 INTERNATIONAL STUDY OF ADULT LITERACY

IALS collected:

- Background assessments
- Literacy assessments

Components

Each IALS country was given a set of model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments. IALS instruments consisted of three parts: (1) a background questionnaire, which collected demographic information about respondents; (2) a set of core literacy tasks, which screened out respondents with very limited literacy skills; and (3) a main booklet of literacy tasks, used to calibrate literacy levels.

Background Questionnaire. The background questionnaire collected information on languages spoken or read; parents' educational attainment and employment; labor force experiences—employment status, recent labor force experiences, and occupation; reading and writing at work and looking for work; participation in adult education classes—courses taken, financial support, purpose; reading and writing in daily life (excluding work or school); family literacy—children's reading habits, the household's access to reading materials, hours spent watching television; and household information—total income and sources of income. The background questionnaire was to be administered in about 20 minutes.

Literacy Assessment—Core Literacy Tasks and Main Literacy Tasks. One hundred and fourteen tasks were grouped into three scales and divided into seven blocks (labeled A through G), which in turn were compiled into seven test booklets (numbered 1 through 7). Each booklet contained three blocks of tasks and was designed to take about 45 minutes to complete. Respondents began the cognitive part of the assessment by performing a set of six "core" tasks. Only those who were able to perform at least two of the six core tasks correctly (93 percent of respondents) were given the full assessment.

Periodicity

The first phase of data collection for IALS was conducted during the autumn of 1994 in Canada, Germany, Ireland, the Netherlands, Poland, Sweden, Switzerland (French and German-speaking cantons), and the United States. Data were collected from a second group of countries or territories—Australia, Flemish-speaking Belgium, Great Britain, New Zealand, and Northern Ireland—in 1995–96. Data were collected from a third group of countries in 1997–98. No second administration is planned.

2. USES OF DATA

IALS was designed to inform both educational and labor market policy and program development

activities in participating countries. The primary objectives of the study were to

- shed light on the relationship between microeconomic variables—such as individual literacy, educational attainment, labor market participation and employment, and macroeconomic issues—such as competitiveness, growth, and restructuring;
- identify subpopulations that are economically and socially disadvantaged by their literacy skill profiles; and
- establish the comparability of assessments of adult literacy.

IALS data provide comparable information about the activities and outcomes of educational systems and institutions in participating countries. Such data can lead to improvements in accountability and policymaking. These data are relevant to policy formation due to the growing political, economic, and cultural ties between countries.

3. KEY CONCEPTS

Some of the key concepts related to the IALS literacy assessment are described below.

Literacy. The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.

Prose Literacy. The ability to read and use texts of varying levels of difficulty that are presented in sentence and paragraph form, including editorials, news stories, poems, and fiction.

Document Literacy. The knowledge and skills required to locate and use information contained in formats such as job applications, payroll forms, transportation schedules, maps, tables, and graphics.

Quantitative Literacy. The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, calculating a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

Literacy Scales. The three scales used to report the results for prose, document, and quantitative literacy. These scales, each ranging from 0 to 500, are based on those established for the Young Adult Literacy Assessment, the DOL's Workplace Literacy Survey,

and the National Adult Literacy Survey. The scores on each scale represent degrees of proficiency along that particular dimension of literacy. The scales make it possible not only to summarize the literacy proficiencies of the total population and of various subpopulations, but also to determine the relative difficulty of the literacy tasks administered in IALS.

The literacy tasks administered in IALS varied widely in terms of materials, content, and task requirements, and thus in difficulty. A careful analysis of the range of tasks along each scale provides clear evidence of an ordered set of information-processing skills and strategies along each scale. To capture this ordering, each scale was divided into five levels that reflect this progression of information-processing skills and strategies: Level 1 (0 to 225), Level 2 (226 to 275), Level 3 (276 to 325), Level 4 (326 to 375), and Level 5 (376 to 500). Level 1 comprised those adults who could consistently succeed with Level 1 literacy tasks but not with Level 2 tasks, as well as those who could not consistently succeed with Level 1 tasks and those who were not literate enough to take the test at all. Adults in Levels 2 through 4 were consistently able to succeed with tasks at their level but not with the next more difficult level of tasks. Adults in Level 5 were consistently able to succeed with Level 5 tasks. The use of three parallel literacy scales makes it possible to profile and compare the various types and levels of literacy demonstrated by adults in different countries and by subgroups within those countries.

4. SURVEY DESIGN

Statistics Canada and ETS, a private testing organization in the United States, coordinated the development and management of IALS. These organizations were assisted by national research teams from the participating countries in developing the survey design. The survey design for the 1994 IALS is described below.

Target Population

The IALS target population was the civilian, noninstitutionalized population ages 16 to 65 in each country; however, countries were also permitted to sample older adults, and several did so. All IALS samples excluded full-time members of the military and people residing in institutions such as prisons, hospitals, and psychiatric facilities.

For the United States, the target population consisted specifically of civilian noninstitutionalized residents ages 16 to 65 in the 50 states and the District of Columbia, excluding members of the armed forces on active duty, those residing outside the United States,

and those with no fixed household address (i.e., the homeless or residents of institutional group quarters, such as prisons and hospitals).

Sample Design

IALS was designed to provide data representative at the national level. Each country that participated in IALS agreed to draw a probability sample that would accurately represent its civilian, noninstitutionalized population ages 16 to 65. The final IALS sample design criteria specified that each country's sample should result in at least 1,000 respondents, the minimum sample size needed to produce reliable literacy proficiency estimates. Given the different sizes of the population of persons ages 16 to 65 in the countries involved, sample sizes varied considerably from country to country (ranging from 1,500 to 8,000 per country), but sample sizes were sufficiently large in all cases to support the estimation of reliable item parameters using Item Response Theory (IRT).

IALS countries were strongly encouraged to select high-quality probability samples because the use of probability designs would make it possible to produce unbiased estimates for individual countries and to compare these estimates across the countries. Because the available data sources and resources were different in each of the participating countries, however, no single sampling methodology was imposed. Each IALS country created its own sample design. All countries used probability sampling for at least some stages of their sample designs, and some used probability sampling for all stages of sampling. Sampling designs were approved by expert review.

The sample for the United States was selected from a sample of individuals in housing units who were completing their final round of interviews for the U.S. Census Bureau's Current Population Survey (CPS) in March, April, May, and June 1994. These housing units were included in the CPS for their initial interviews in December 1992 and January, February, and March 1993. The CPS is a large-scale continuous household survey of the civilian noninstitutionalized population age 15 and over. The frame for the CPS consisted of 1990 decennial census files, which are continually updated for new residential construction and are adjusted for undercount, births, deaths, immigration, emigration, and changes in the armed forces.

The CPS sample is selected using a stratified multistage design. Housing units that existed at the time of the 1990 population census were sampled from the census list of addresses. Housing units that did not exist at that time were sampled from lists of new

construction, when available, and otherwise by area sampling methods. Occupants of housing units that came into existence between the time of the CPS sample selection and the time of the IALS fieldwork had no chance of being selected for IALS.

The IALS sample was confined to 60 of the 729 CPS primary sampling units (PSUs). Within these 60 PSUs, all persons 16 to 65 years of age in the sampled housing units were classified into 20 cells defined by race/ethnicity and education. Within each cell, persons were selected for IALS with probability proportional to their CPS weights, with the aim of producing an equal probability sample of persons within cells. A total of 4,901 persons were selected for IALS. IALS interviews were conducted in October and November 1994.

Assessment Design

The success of IALS depended on the development and standardized application of a common set of survey instruments. The test framework explicitly followed the precedent set by the National Adult Literacy Survey, basing the test on U.S. definitions of literacy along three dimensions—prose literacy, document literacy, and quantitative literacy—but extending the instruments into an international context. Study managers from each participating country were encouraged to submit materials such as news articles and documents that could be used to create tasks with the goal of building a new pool of literacy tasks that could be linked to established scales. IALS field tested 175 tasks and identified 114 that were valid across cultures. Approximately half of these tasks were based on materials from outside North America. (However, each respondent was administered only a fraction of the pool of tasks, using a variant of matrix sampling.)

Each IALS country was given a set of model administration manuals and survey instruments as well as graphic files containing the pool of IALS literacy items with instructions to modify each item by translating the English text to its own language without altering the graphic representation. Certain rules governed the item modification process. For instance, some items required respondents to perform a task that was facilitated by the use of keywords. The keyword in the question might be identical to, similar but not exactly the same as, or a synonym of the word used in the body of the item, or respondents might be asked to choose among multiple keywords in the body of the item, only one of which was correct. Countries were required to preserve these conceptual associations during the translation process. Particular conventions used in the items—for example, currency units, date formats, and decimal delimiters—were adapted as appropriate for each country.

To ensure that the adaptation process did not compromise the psychometric integrity of the items, each country's test booklets were carefully reviewed for errors of adaptation. Countries were required to correct all errors found. However, this review was imperfect in two important respects. First, it is clear that countries chose not to incorporate a number of changes that were identified during the course of the review, believing that they "knew better." Second, the availability of empirical data from the study has permitted the identification of several additional sources of task and item difficulty that were not included in the original framework, which was based on research by Irwin Kirsch of ETS and Peter Mosenthal of Syracuse University. (See 1990 publication, "[Exploring Document Literacy: Variables Underlying the Performance of Young Adults](#)," by I.S. Kirsch and P.B. Mosenthal in *Reading Research Quarterly* 25: 5–30.) Item adaptation guidelines and item review procedures associated with subsequent rounds of IALS data collection were adapted to reflect this additional information.

The model background questionnaires contained two sets of questions: mandatory questions, which all countries were required to include; and optional questions, which were recommended but not required. Countries were not required to field literal translations of the mandatory questions, but were asked to respect the conceptual intent of each question in adapting it for use. Countries were permitted to add questions to their background questionnaires if the additional burden on respondents would not reduce response rates. Statistics Canada reviewed all background questionnaires (except Sweden's) before the pilot survey and offered comments and suggestions to each country.

Data Collection and Processing

IALS data for the first round of countries were collected through in-person household interviews in the fall of 1994. Each country mapped its national dataset into a highly structured, standardized record layout that it sent to Statistics Canada. Further description follows.

Reference dates. Respondents answered questions about jobs they may have held in the 12 months before the survey was administered.

Data collection. Statistics Canada and ETS coordinated the development and management of IALS. Participating countries were given model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments and for handling nonresponse codings.

Countries were permitted to adapt these models to their own national data collection systems, but they were required to retain a number of key features: (1) respondents were to complete the core and main test booklets alone, in their homes, without help from another person or from a calculator; (2) respondents were not to be given monetary incentives for participating; (3) despite the prohibition on monetary incentives, interviewers were provided with procedures to maximize the number of completed background questionnaires and were to use a common set of coding specifications to deal with nonresponse. This last requirement was critical. Because noncompletion of the core and main task booklets was correlated with ability, background information about nonrespondents was needed in order to impute cognitive data for these persons.

IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. All countries participating in IALS instructed interviewers to make callbacks at households that were difficult to contact.

In general, the survey was carried out in the national language. In Canada, respondents were given a choice of English or French, and in Switzerland, samples drawn from French-speaking and German-speaking cantons were required to respond in those respective languages. When respondents could not speak the designated language, attempts were made to complete the background questionnaire so that their literacy level could be estimated and the possibility of distorted results would be reduced. In the United States, the test was given in English, but a Spanish version of the background questionnaire and bilingual interviewers were available to assist individuals whose native language was not English.

Survey respondents spent approximately 20 minutes answering a common set of background questions concerning their demographic characteristics, educational experiences, labor market experiences, and literacy-related activities. Responses to these background questions made it possible to summarize the survey results using an array of descriptive variables, and also increased the accuracy of the proficiency estimates for various subpopulations. After answering the background questions, the remainder of respondents' time was spent completing a booklet of literacy tasks designed to measure their prose, document, and quantitative skills. Most of these tasks were open-ended, requiring respondents to provide a written answer.

In the United States, the IALS interview period was from October to November 1994. IALS was conducted by 149 Census Bureau interviewers. All of them had at least 5 days of interviewer training. They were given a one-day training on IALS and were provided with substantial training and reference materials based on the Canadian training package. They also performed a day of field training under the supervision of a regional office supervisor. Each interviewer had an average workload of 33 interviews, and the average number of response interviews per interviewer was 21. They were supervised by six regional supervisors who reviewed and commented on their work.

Before data collection, a letter was sent to the selected addresses describing the upcoming survey. The survey was limited to 90 minutes. If a respondent took more than 20 minutes per block, the interviewer was instructed to move the respondent on to the next block.

Data processing. As a condition of their participation in IALS, countries were required to capture and process their files using procedures that ensured logical consistency and acceptable levels of data capture error. Specifically, countries were advised to conduct complete verification of the captured scores (i.e., enter each record twice) in order to minimize error rates. One hundred percent keystroke validation was needed. Specific details about scoring are provided in a separate section below.

To create a workable comparative analysis, each IALS country was required to map its national dataset into a highly structured, standardized record layout. In addition to specifying the position, format, and length of each field, this International Record Layout included a description of each variable and indicated the categories and codes to be provided for that variable. Upon receiving a country's file, Statistics Canada performed a series of range checks to ensure compliance to the prescribed format. When anomalies were detected, countries corrected the problems and submitted new files. Statistics Canada did not, however, perform any logic or flow edits, as it was assumed that participating countries performed this step themselves.

Editing. Most countries followed IALS guidelines, verifying 100 percent of their data capture operation. The two countries that did not comply with this recommendation conducted sample verifications, one country at 20 percent and the other at 10 percent. Each country coded and edited its own data, mapping its national dataset into the detailed International Record Layout, which included a description of each variable and indicated the categories and codes to be provided

for that variable. Industry, occupation, and education were coded using the standard international coding schemes: the International Standard Industrial Classification (ISIC), the International Standard Classification of Occupations (ISCO), and the International Standard Classification of Education (ISCED). Coding schemes were provided for open-ended items; the coding schemes came with specific instructions so that coding error could be contained to acceptable levels.

Scoring. Respondents' literacy proficiencies were estimated based on their performance on the cognitive tasks administered in the assessment. Because the open-ended items used in IALS elicited a large variety of responses, responses had to be grouped in order to summarize the performance results. As they were scored, responses to IALS open-ended items were classified as correct, incorrect, or omitted. The models employed to estimate ability and difficulty were predicated on the assumption that the scoring rubrics developed for the assessment were applied in a consistent fashion within and between countries. To reinforce the importance of consistent scoring, a meeting of national study managers and chief scorers was held prior to the commencement of scoring for the main study. The group spent 2 days reviewing the scoring rubrics for all the survey items. Where this review uncovered ambiguities and situations not covered by the guides, clarifications were agreed to collectively, and these clarifications were then incorporated into the final rubrics. To provide ongoing support during the scoring process, Statistics Canada and ETS maintained a joint scoring hotline. Any scoring problems encountered by chief scorers were resolved by this group, and decisions were forwarded to all national study managers. Study managers conducted intensive scoring training using the scoring manual and discussed unusual responses with scorers. They also offered additional training to some scorers, as needed, to raise their accuracy to the level achieved by other scorers.

To maintain coding quality within acceptable levels of error, each country undertook to rescore a minimum of 10 percent of all assessments. Where significant problems were encountered, larger samples of a particular scorer's work were to be reviewed and, where necessary, their entire assignments rescored. Countries were not required to resolve contradictory scores in the main survey (as they had been in the pilot), since outgoing agreement rates were far above minimum acceptable tolerances.

Since there could still be significant differences in the consistency of scoring between countries, countries

agreed to exchange at least 300 randomly selected booklets with another country sharing the same test language. In all cases where serious discrepancies were identified, countries were required to rescore entire items or discrepant code pairs.

Intra-country rescoring. A variable sampling ratio procedure was set up to monitor scoring accuracy. At the beginning of scoring, almost all responses were rescored to identify inaccurate scorers and to detect unique or difficult responses that were not covered in the scoring manual. After a satisfactory level of accuracy was achieved, the rescoring ratio was dropped to a maintenance level to monitor the accuracy of all scorers. Average agreements were calculated across all items. Precautions were taken to ensure that the first and second scores were truly independent.

Inter-country rescoring. To determine intercountry scoring reliabilities for each item, the responses of a subset of examinees were scored by two separate groups. Usually, these scoring groups were from different countries. Intercountry score reliabilities were calculated by Statistics Canada, and then evaluated by ETS. Based on the evaluation, every country was required to introduce a few minor changes in scoring procedures. In some cases, ambiguous instructions in the scoring manual were found to be causing erroneous interpretations and therefore lower reliabilities.

Using the intercountry score reliabilities, researchers could identify poorly constructed items, ambiguous scoring criteria, erroneous translations of items or scoring criteria, erroneous printing of items or scoring criteria, scorer inaccuracies, and, most important, situations in which one country consistently scored differently from another. In the latter circumstance, scorers in one country may consistently rate a certain response as being correct while those in another country score the same response as incorrect. ETS and Statistics Canada examined scoring carefully to identify situations in which scorers in one country were consistently rating a certain response as being correct while those in another country were scoring the same response as incorrect. Where a systematic error was identified in a particular country, the original scores for that item were corrected for the entire sample.

Estimation Methods

Weighting was used in the 1994 IALS to adjust for sampling and nonresponse. Responses to the literacy tasks were scored using IRT scaling. A multiple imputation procedure based on plausible values methodology was used to estimate the literacy proficiencies of individuals who completed literacy tasks.

Weighting. IALS countries used different methods for weighting their samples. Countries with known probabilities of selection could calculate a base weight using the probability of selection. To adjust for unit nonresponse, all countries poststratified their data to known population counts, and a comparison of the distribution of the age and sex characteristics of the actual and weighted samples indicates that the samples were comparable to the overall populations of IALS countries. Another commonly used approach was to weight survey data to adjust the rough estimates produced by the sample to match known population counts from sources external to IALS. This “benchmarking” procedure assumes that the characteristics of nonrespondents are similar to those of respondents. It is most effective when the variables used for benchmarking are strongly correlated with the characteristic of interest—in this case, literacy levels. For IALS, the key benchmarking variables were age, employment status, and education. All of the IALS countries benchmarked to at least one of these variables. The United States used education.

Weights for the U.S. IALS sample included two components. The first assigned weights to CPS respondents, and the second assigned weights to IALS respondents.

The CPS weighting scheme was a complex one involving three components: basic weighting, noninterview adjustment, and ratio adjustment. The basic weighting compensated for unequal selection probabilities. The noninterview adjustment compensated for nonresponse within weighting cells created by clusters of PSUs of similar size; Metropolitan Statistical Area (MSA) clusters were subdivided into central city areas, and the balance of the MSA and non-MSA clusters were divided into urban and rural areas. The ratio adjustment made the weighted sample distributions conform to known distributions on such characteristics as age, race, Hispanic origin, sex, and residence.

The weights of persons sampled for IALS were adjusted to compensate for the use of the four rotation groups, the sampling of the 60 PSUs, and the sampling of persons within the 60 PSUs. The IALS noninterview adjustment compensated for sampled persons for whom no information was obtained because they were absent, refused to participate, had a short-term illness, had moved, or had experienced an unusual circumstance that prevented them from being interviewed. Finally, the IALS ratio adjustment ensured that the weighted sample distributions across a number of education groups conformed to March 1994 CPS estimates of these numbers.

Scaling. The scaling model used in IALS was the two-parameter logistic model based on IRT.

Items developed for IALS were based on the framework used in three previous large-scale assessments: the Young Adult Literacy Assessment, the DOL survey, and the National Adult Literacy Survey. As a result, IALS items shared the same characteristics as the items in these earlier surveys. The English versions of IALS items were reviewed and tested to determine whether they fit into the literacy scales in accordance with the theory and whether they were consistent with the National Adult Literacy Survey data. Quality control procedures for item translation, scoring, and scaling followed the same procedures used in the National Adult Literacy Survey and extended the methods used in other international studies.

Identical item calibration procedures were carried out separately for each of the three literacy scales: prose, document, and quantitative literacy. Using a modified version of Mislevy and Bock’s 1982 BILOG computer program—see *BILOG: Item analysis and test scoring with binary logistic models*, Scientific Software—the two-parameter logistic IRT model was fit to each item using sample weights. BILOG procedures are based on an extension of the marginal-maximum-likelihood approach described by Bock and Aitkin in their 1981 *Psychometrika* article, “Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm.”

Most of the items administered in IALS were successful from a psychometric standpoint. However, despite stringent efforts at quality control, some of the assessment items did not meet the criteria for inclusion in the final tabulation of results. Specifically, in carrying out the IRT modeling used to create the three literacy scales, researchers found that a number of assessment items had significantly different item parameters across IALS countries.

Imputation. A respondent had to complete the background questionnaire, pass the core block of literacy tasks, and attempt at least five tasks per literacy scale in order for researchers to be able to estimate his or her literacy skills directly. Literacy proficiency data were imputed for individuals who failed or refused to perform the core literacy tasks and for those who passed the core block but did not attempt at least five tasks per literacy scale. Because the model used to impute literacy estimates for nonrespondents relied on a full set of responses to the background questions, IALS countries were instructed to obtain at least a background questionnaire from sampled individuals.

IALS countries were also given a detailed nonresponse classification to use in the survey.

Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Special procedures were used to impute missing cognitive data.

Literary proficiency estimation (plausible values). A multiple imputation procedure based on plausible values methodology was used to estimate respondents' literacy proficiency in the 1994 IALS. When a sampled individual decided to stop the assessment, the interviewer used a standardized nonresponse coding procedure to record the reason why the person was stopping. This information was used to classify nonrespondents into two groups: (1) those who stopped the assessment for literacy-related reasons (e.g., language difficulty, mental disability, or reading difficulty not related to a physical disability); and (2) those who stopped for reasons unrelated to literacy (e.g., physical disability or refusal). About 45 percent of the individuals did not complete the assessment for reasons related to their literacy skills; the other respondents gave no reason for stopping or gave reasons unrelated to their literacy.

When individuals cited a literacy-related reason for not completing the cognitive items, it implies that they were unable to respond to the items. On the other hand, citing reasons unrelated to literacy implies nothing about a person's literacy proficiency. Based on these interpretations, IALS adapted a procedure originally developed for the National Adult Literacy Survey to treat cases in which an individual responded to fewer than five items per literacy scale, as follows: (1) if the individual cited a literacy-related reason for not completing the assessment, then all consecutively missing responses at the end of the block of items were treated as wrong; and (2) if the individual cited reasons unrelated to literacy for not completing the assessment, then all consecutively missing responses at the end of a block were treated as "not reached."

Proficiency values were estimated based on respondents' answers to the background questions and the cognitive items. As an intermediate step, the functional relationship between these two sets of information was calculated, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency was calculated from a posterior distribution that was the multiple of two functions: a conditional distribution of proficiency, given responses to the background questions; and a likelihood function of proficiency, given responses to the cognitive items.

Recent Changes

Since IALS was a one-time assessment, there are no changes to report.

Future Plans

There are no plans to conduct IALS again. However, a new survey, the Adult Literacy and Lifeskills Survey (ALL), was administered in 2003 (see [ALL chapter](#)). The aspects of this survey that address literacy were built on methodologies used in IALS.

5. DATA QUALITY AND COMPARABILITY

The literacy tasks contained in IALS and the adults asked to participate in the survey were samples drawn from their respective universes. As such, they were subject to some measurable degree of uncertainty. IALS implemented procedures to minimize both sampling and nonsampling errors. The IALS sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. Scientific procedures employed in the study design and the scaling of literacy tasks permitted a high degree of confidence in the resulting estimates of task difficulty. Quality control activities continued during interviewer training, data collection, and processing of the survey data.

In addition, special evaluation studies were conducted to examine issues related to the quality of IALS. These studies included (1) an external evaluation of IALS methodology; (2) an examination of how similar or different the sampled persons were from the overall population; (3) an evaluation of the extent to which the literacy levels of the population in the database for each nation were predictable based on demographic characteristics; (4) an examination of the assumption of unidimensionality; and (5) an evaluation of the construct validity of the adult literacy scales.

Sampling Error

Because IALS employed probability sampling, the results were subject to sampling error. Although small, this error was higher in IALS than in most studies because the cost of surveying adults in their homes is so high. Most countries simply could not afford large sample sizes.

Each country provided a set of replicate weights for use in a jackknife variance estimation procedure.

There were three situations in which nonprobability-based sampling methods were used: France and Germany used "random route" procedures for selecting

households into their samples, and Switzerland used an alphabetic sort to select one member of each household. However, based on the available evidence, it is not believed that these practices introduced significant bias into the survey estimates.

In 1998, the U.K. Office of National Statistics coordinated the European Adult Literacy Review, a split-sample survey intended, in part, to measure the effects of sampling methods on the IALS results. This follow-up survey compared an IALS sample design with an alternative, standardized “best practice” design. Although certain differences were noted between the two samples, the IALS sample design was not confirmed to be inferior to the “best practice” design.

Nonsampling Error

The key sources of nonsampling error in the 1994 IALS were differential coverage across countries and nonresponse bias, which occurred when different groups of sampled individuals failed to participate in the survey. Other potential sources of nonsampling error included deviations from prescribed data collection procedures and errors of logic that resulted from mapping idiosyncratic national data into a rigid international format. Scoring error, associated with scoring open-ended tasks reliably within and between countries, also occurred. Finally, because IALS data were collected and processed independently by the various countries, the study was subject to uneven levels of commonplace data capture, data processing, and coding errors.

Three studies were conducted to examine the possibility of nonresponse bias. Because the sampling frames for Canada and the United States contained information about the characteristics of sampled individuals, it was possible to compare the characteristics of respondents and nonrespondents, particularly with respect to literacy skill profiles. The Swedish National Study Team also commissioned a nonresponse follow-up study.

Coverage error. The design specifications for IALS stated that in each country the study should cover the civilian, noninstitutionalized population ages 16 to 65. It is the usual practice to exclude the institutional population from national surveys because of the difficulties in conducting interviews in institutional settings. Similarly, it is not uncommon to exclude certain other parts of a country’s population that pose difficult survey problems (e.g., persons living in sparsely populated areas). The intended coverage of the surveys generally conformed well to the design specifications: each of the IALS countries attained a

high level of population coverage, ranging from a low of 89 percent in Switzerland to a high of 99 percent in the Netherlands and Poland. However, it should be noted that actual coverage is generally lower than the intended coverage because of deficiencies in sampling frames and sampling frame construction (e.g., failures to list some households and some adults within listed households). In the United States, for example, comparing population sizes estimated from the survey with external benchmark figures suggests that the overall coverage rate for the CPS (the survey from which the IALS sample was selected) is about 93 percent, but that it is much lower for certain population subgroups (particularly young Black male adults).

Nonresponse error. For IALS, several procedures were developed to reduce biases due to nonresponse, based on how much of the survey the respondent completed.

Unit nonresponse. The definition of a respondent for IALS was a person who partially or fully completed the background questionnaire. Unweighted response rates varied considerably from country to country, ranging from a high of 69 percent (Canada, Germany) to a low of 45 percent (the Netherlands), with four countries in the 55–60 percent range.

In the United States, which had a response rate of 60 percent, nonresponse to IALS occurred for two reasons: (1) some individuals did not respond to the CPS; and (2) some of the CPS respondents selected for IALS did not respond to the IALS instruments. In any given month, nonresponse to the CPS is typically quite low, around 4 to 5 percent. Its magnitude in the expiring rotation groups employed for IALS selection is not known. About half of the CPS nonresponse is caused by refusals to participate, while the remainder is caused by temporary absences, other failures to contact individuals, the inability of individuals contacted to respond, and unavailability for other reasons.

A sizable proportion of the nonresponse to the IALS background questionnaire was attributable to persons who had moved. For budgetary reasons, it was decided that persons who were not living at the CPS addresses at the time of the IALS interviews would not be contacted. This decision had a notable effect on the sample of students, who are sampled in dormitories and other housing units in the CPS only if they do not officially reside at their parents’ homes. Those who reside at their parents’ homes are included in the CPS at that address, but because most of these students were away at college during the IALS interview period (October to November 1994), they could not respond to IALS.

The high level of nonresponse for college students could cause a downward bias in the literacy skill-level estimates. This group represents only a small proportion of the U.S. population, however, so the potential bias is likely to be quite small. Furthermore, a comparison of IALS results to the U.S. National Adult Literacy Survey data discounts this as a major source of bias.

Item nonresponse. The weighted percentage of omitted responses for the U.S. IALS sample ranged from 0 to 18 percent.

Not-reached responses were classified into two groups: nonparticipation immediately or shortly after the background information was collected; and premature withdrawal from the assessment after a few cognitive items were attempted. The first type of not-reached response varied a great deal across countries according to the frames from which the samples were selected. The second type of not-reached response was due to quitting the assessment early, resulting in incomplete cognitive data. Not-reached items were treated as if they provided no information about the respondent's proficiency, so they were not included in the calculation of likelihood functions for individual respondents. Therefore, not-reached responses had no direct impact on the proficiency estimation for subpopulations. The impact of not-reached responses on the proficiency distributions was mediated through the subpopulation weights.

Measurement error. Assessment tasks were selected to ensure that, among population subgroups, each literacy domain (prose, document, and quantitative) was well covered in terms of difficulty, stimuli type, and content domain. The IALS item pool was developed collectively by participating countries. Items were subjected to a detailed expert analysis at ETS and vetted by participating countries to ensure that the items were culturally appropriate and broadly representative of the population being tested. For each country, experts who were fluent in both English and the language of the test reviewed the items and identified ones that had been improperly adapted. Countries were asked to correct problems detected during this review process. To ensure that all of the final survey items had a high probability of functioning well, and to familiarize participants with the unusual operational requirements involved in data collection, each country was required to conduct a pilot survey. Although the pilot surveys were small and typically were not based strictly on probability samples, the information they generated enabled ETS to reject items, to suggest modifications to a few items, and to choose good items for the final assessment. ETS's

analysis of the pilot survey data and recommendations for the final test design were presented to and approved by participating countries.

Data Comparability

While most countries closely followed the data collection guidelines provided, some did deviate from the instructions. First, two countries (Sweden and Germany) offered participation incentives to individuals sampled for their survey. The incentive paid was trivial, however, and it is unlikely that this practice distorted the data. Second, the doorstep introduction provided to respondents differed somewhat from country to country. Three countries (Germany, Switzerland, and Poland) presented the literacy test booklets as a review of the quality of published documents rather than as an assessment of the respondent's literacy skills. A review of these practices suggested that they were intended to reduce response bias and were warranted by cultural differences in respondents' attitudes toward being tested. Third, there were differences across the countries in the way in which interviewers were paid. No guidelines were provided on this subject, and the study teams therefore decided what would work best in their respective countries. Fourth, several countries adopted field procedures that undermined the objective of obtaining completed background questionnaires for an overwhelming majority of selected respondents.

This project was designed to produce data comparable across cultures and languages. After one of the countries in the first round raised concerns about the international comparability of the survey data, Statistics Canada decided that the IALS methodology should be subjected to an external evaluation. In the judgment of the expert reviewers, the considerable efforts that were made to develop standardized survey instruments for the different nations and languages were successful, and the data obtained from them should be broadly comparable.

However, the standardization of procedures with regard to other aspects of survey methodology was not achieved to the extent desired, resulting in several weaknesses. Nonresponse proved to be a particular weakness, with generally very high nonresponse rates and variation in nonresponse adjustment procedures across countries. For some countries the sample design was problematic, resulting in some unknown biases. The data collection and its supervision differed between participating countries, and some clear weaknesses were evident for some countries. The reviewers felt that the variation in survey execution across countries was so large that they recommended against publication of comparisons of overall national

literacy levels. They did, however, despite the methodological weaknesses, recommend that the survey results be published. They felt that the instruments developed for measuring adult literacy constituted an important advance, and the results obtained for the instruments in the first round of IALS were a valuable contribution to the field. They recommended that the survey report focus on analyses of the correlates of literacy (e.g., education, occupation, and age) and the comparison of these correlates across countries. Although these analyses might also be distorted by methodological problems, they believed that the analyses were likely to be less affected by these problems than were the overall literacy levels.

6. CONTACT INFORMATION

For content information on IALS, contact:

Eugene Owen
Phone: (202) 502-7422
E-mail: eugene.owen@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

Murray, T.S., Kirsch, I.S. and Jenkins, L.B. (eds.). (1997). *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey* (NCES 98-053). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.