

Adult Literacy and Lifeskills Survey (ALL)

Website: <http://nces.ed.gov/surveys/all/>

1. OVERVIEW

The Adult Literacy and Lifeskills Survey (ALL) is an international comparative study designed to provide participating countries, including the United States, with information about the skills of their adult populations ages 16 to 65. The development and management of the study were coordinated by Statistics Canada and the Educational Testing Service (ETS) in collaboration with the National Center for Education Statistics (NCES) of the U.S. Department of Education; the Organization for Economic Cooperation and Development (OECD); the Regional Office of Education for Latin America and the Caribbean (OREALC); and the Institute for Statistics (UIS) of the United Nations Educational, Scientific, and Cultural Organization (UNESCO).

ALL measured the literacy and numeracy skills of a nationally representative sample from each participating country. On a pilot basis, ALL also measured adults' problem-solving skills and gathered information on their familiarity with information and communication technology (ICT). ALL builds on the foundation of earlier studies of adult literacy. Chief among these earlier studies is the International Adult Literacy Survey (IALS), which was conducted in three phases (1994, 1996, and 1998) in 20 nations, including the United States. The following six countries participated in ALL: Italy, Norway, Switzerland, Bermuda, Canada, and the United States.

Purpose

To (1) profile and compare the literacy skills in adult populations; (2) profile and compare the level and distribution of directly assessed numeracy skills among adult populations in participating countries; (3) profile and compare the level and distribution of problem-solving skills among the adult populations of the countries surveyed; and (4) collect comparable data on participation in formal adult education.

Components

Each ALL country was given a set of model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments. ALL 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 general participant information (such as sex, age, race/ethnicity, education level, and labor force status) and posed more targeted questions related to literacy practices, familiarity with ICT, education coursetaking, and health.

ADULT LITERACY AND LIFESKILLS SURVEY

ALL collected:

- Background assessments
- Literacy assessments in prose literacy, document literacy, numeracy, and problem-solving domains

Literacy Assessment.

Core literacy tasks. The core literacy tasks were presented to respondents once they had completed the background questionnaire. The booklet for the core literacy tasks contained six simple tasks. Only those who answered at least two of the core tasks correctly were given the full assessment.

Main literacy tasks. The main literacy tasks for the ALL psychometric assessment consisted of tasks in prose literacy, document literacy, numeracy, and problem-solving domains. The assessment included four 30-minute blocks of literacy items (i.e., prose and document literacy), two 30-minute blocks of numeracy items, and two 30-minute blocks of problem-solving items. A four-domain ALL assessment was implemented in Bermuda, Canada, Italy, Norway, and the French- and German-language regions of Switzerland. The United States and the Italian-language region of Switzerland carried out a three-domain ALL assessment that excluded the problem-solving domain. The blocks of assessment items were organized into 28 task booklets in the case of the four-domain assessment and into 18 task booklets for the three-domain assessment. The assessment blocks were distributed to the task booklets according to a balanced incomplete block (BIB) design whereby each task booklet contained two blocks of items.

Periodicity

ALL was conducted between the fall of 2003 and early spring 2004. In the United States, data collection for the main study took place between January and June 2003.

2. USES OF DATA

ALL sought to provide researchers with information on skill gain and loss in the adult population. This was achieved through the measurement of prose and document literacy. Furthermore, the study extended the range of skills measured by adding tasks for problem-solving, numeracy, and ICT skills. This allows researchers to examine the profiles of important foundation skills. The study makes it possible to explore the interrelationships among skill domains as well as their links to major antecedents and outcomes, such as the quantity and quality of initial education and the impact of skills on employability, wages, and health.

In addition, information from ALL addresses questions such as the following:

- What is the distribution of literacy and numeracy skills among American adults? How do these skill distributions compare to those of other countries?
- What is the relationship between these literacy skills and the economic, social, and personal characteristics of individuals? For example: Do different age or linguistic groups manifest different skill levels? Do males and females perform differently? At what kinds of jobs do people at various literacy levels work? What wages do they earn? How do adults who have completed different levels of education perform?
- What is the relationship between these skills and the economic and social characteristics of nations? For example, how do the skills of the adult labor force of a country match up with areas of the economy that are growing?

3. KEY CONCEPTS

Four skill domains are conceptualized in ALL: prose literacy, document literacy, numeracy, and problem solving. Two of them, namely, prose and document literacy, are defined and measured in the same manner as in IALS (see [IALS chapter](#)). Numeracy and problem solving are new domains.

Prose literacy. The knowledge and skills needed to understand and use information from texts, including editorials, news stories, brochures, and instruction manuals.

Document literacy. The knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and charts.

Numeracy. The knowledge and skills required to effectively manage the mathematical demands of diverse situations.

Problem solving. Problem solving involves goal-directed thinking and action in situations for which no routine solution procedure is available. The problem solver has a more or less well-defined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a problem. The understanding of the problem situation and its step-by-step transformation based on planning and reasoning constitute the process of problem solving.

Literacy scale. For each skill assessment domain, proficiency is denoted on a scale ranging from 0 to 500 points. Each score denotes a point at which a person has an 80 percent chance of successfully completing tasks that are associated with a similar level of difficulty. For the prose and document literacy domains as well as the numeracy domain, experts defined five broad levels of difficulty, each corresponding to a range of scores. For the problem-solving domain, experts defined four broad levels of difficulty.

4. SURVEY DESIGN

Each participating country was required to design and implement the Adult Literacy and Lifeskills Survey according to specified guidelines and standards. These ALL standards established the minimum survey design and implementation requirements for the following project areas: survey planning; target population; method of data collection; sample frame; sample design; sample selection; literacy assessment design; background questionnaire; task booklets; instrument requirements to facilitate data processing; data collection; respondent contact strategy; response rate strategy; interviewer hiring, training, and supervision; data capture; coding and scoring; data file format and editing; weighting; estimation; confidentiality; survey documentation; and pilot survey.

Target Population

Each participating country designed a sample to be representative of its civilian noninstitutionalized population ages 16 to 65 (inclusive). Countries were also at liberty to include adults over the age of 65 in the sample provided that a minimum suggested sample size requirement was satisfied for the 16 to 65 age group. Canada opted to include in its target population adults over the age of 65. All of the remaining countries restricted the target population to the 16 to 65 age group. Exclusions from the target population for practical operational reasons were acceptable provided a country's survey population did not differ from the target population by more than 5 percent (i.e., provided that the total number of exclusions from the target population due to undercoverage was not more than 5 percent of the target population). All countries indicate that this 5 percent requirement was satisfied. Each country chose or developed a sample frame to cover the target population.

Sample Design

Each participating country was required to use a probability sample representative of the national population ages 16 to 65. A sample size of 5,400 completed cases in each official language was recommended for each country that was implementing the full ALL psychometric assessment (i.e., comprising the prose literacy, document literacy, numeracy, and problem-solving domains). A sample size of 3,420 complete cases in each official language was recommended if the problem-solving domain was excluded from the ALL assessment.

The available sampling frames and resources varied from one country to another. Therefore, the particular probability sample design to be used was left to the discretion of each country. Each country's proposed sample design was reviewed by Statistics Canada to ensure that the sample design standards and guidelines were satisfied.

A stratified multistage probability sample design was employed in the United States. The first stage of sampling consisted of selecting a sample of 60 primary sampling units (PSUs) from a total of 1,880 PSUs that were formed using a single county or a group of contiguous counties, depending on the population size and the area covered by a county or counties. The PSUs were stratified on the basis of the social and economic characteristics of the population, as reported in the 2000 census. The following characteristics were used to stratify the PSUs: region of the country, Metropolitan Statistical Area (MSA), population size, percentage of African-American residents, percentage of Hispanic residents, and per capita income. The largest PSUs in terms of population size were included in the sample with certainty. For the remaining PSUs, one PSU per stratum was selected with probability proportional to the population size.

At the second sampling stage, a total of 505 geographic segments were systematically selected with probability proportional to population size from the sampled PSUs. Segments consist of area blocks (as defined by the 2000 census) or combinations of two or more nearby blocks. They were formed to satisfy criteria based on population size and geographic proximity. The third stage of sampling involved the listing of the dwellings in the selected segments and the subsequent selection of a random sample of dwellings. An equal number of dwellings was selected from each sampled segment. At the fourth and final stage of sampling, one eligible person was randomly selected within households with fewer than four eligible adults. In households with four or

more eligible persons, two adults were randomly selected.

Assessment Design

A BIB assessment design was used to measure the skill domains. The BIB design comprised a set of assessment tasks organized into smaller sets of tasks, or blocks. Each block contained assessment items from one of the skill domains and covered a wide range of difficulty (i.e., from easy to difficult). The blocks of items were organized into task booklets according to a BIB design. Individual respondents were not required to take the entire set of tasks. Instead, each respondent was randomly administered one of the task booklets.

ALL assessment. The ALL psychometric assessment consisted of the prose literacy, document literacy, numeracy, and problem-solving domains. The assessment included four 30-minute blocks of literacy items (i.e., prose and document literacy), two 30-minute blocks of numeracy items, and two 30-minute blocks of problem-solving items. A four-domain ALL assessment was implemented in Bermuda, Canada, Italy, Norway, and the French- and German-language regions of Switzerland. The United States and the Italian-language region of Switzerland carried out a three-domain ALL assessment that excluded the problem-solving domain.

The blocks of assessment items were organized into 28 task booklets in the four-domain assessment and into 18 task booklets in the three-domain assessment. The assessment blocks were distributed to the task booklets according to a BIB design whereby each task booklet contained two blocks of items. The task booklets were randomly distributed among the selected sample. In addition, the data collection activity was closely monitored in order to obtain approximately the same number of complete cases for each task booklet, except for two-task booklets in the three-domain assessment containing only numeracy items, which required a larger number of complete cases.

Data Collection and Processing

The data collection for the ALL project took place between the fall of 2003 and early spring 2004, depending on the country. However, in the United States, data collection for the main study took place between January and June 2003. In the United States, a nationally representative sample of 3,420 adults ages 16 to 65 participated in ALL. Trained interviewers administered approximately 45 minutes of background questions and 60 minutes of assessment items to participants in their homes.

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

Data collection. The ALL survey design combined educational testing techniques with those of household survey research to measure literacy and provide the information necessary to make these measures meaningful. The respondents were first asked a series of questions to obtain background and demographic information on educational attainment, literacy practices at home and at work, labor force information, ICT use, adult education participation, and literacy self-assessment. Once the background questionnaire had been completed, the interviewer presented a booklet containing six simple tasks (the core tasks). Respondents who passed the core tasks were given a much larger variety of tasks, drawn from a pool of items grouped into blocks; each booklet contained two blocks that represented about 45 items. No time limit was imposed on respondents, and they were urged to try each item in their booklet. Respondents were given the maximum leeway to demonstrate their skill levels, even if their measured skills were minimal.

To ensure high-quality data, ALL guidelines specified that each country should work with a reputable data collection agency or firm, preferably one with its own professional, experienced interviewers. The interviews were to be conducted in the home in a neutral, nonpressured manner. Interviewer training and supervision was to be provided that emphasized the selection of one person per household (if applicable), the selection of one of the 28 main task booklets (if applicable), the scoring of the core task booklet, and the assignment of status codes. Finally, the interviewers' work was to be supervised by the use of quality checks—frequent quality checks at the beginning of the data collection and fewer quality checks throughout the remainder of the data collection—and by having help available to interviewers during entire the data collection period.

Several precautions were taken against nonresponse bias. Interviewers were specifically instructed to return several times to nonrespondent households in order to obtain as many responses as possible. In addition, all countries were asked to ensure that the address information provided to interviewers was as complete as possible in order to reduce potential household identification problems. Countries were asked to complete a debriefing questionnaire after the study in order to demonstrate that the guidelines had been followed, as well as to identify any collection problems they had encountered.

The United States administered the survey only in English. It used 106 interviewers during the data collection process, assigning approximately 64 cases to each interviewer. Professional interviewers were used to conduct the survey, although approximately one-quarter of the interviewers had no previous survey experience.

Data processing. As a condition of their participation in ALL, 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. Because the process of accurately capturing the task scores is essential to high data quality, 100 percent keystroke verification was required.

Each country was also responsible for coding industry, occupation, and education using standard coding schemes, such as 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 by Statistics Canada for all open-ended items, and countries were given specific instructions about the coding of such items.

In order to facilitate comparability in data analysis, each ALL 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, the 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; flow and consistency edits were also run on the file. When anomalies were detected, countries were notified of the problem and were asked to submit cleaned files.

Scoring. Persons in each country charged with scoring received intense training, using the ALL scoring manual, in scoring responses to the open-ended items. They were also provided a tool for capturing closed format questions. To aid in maintaining scoring accuracy and comparability between countries, ALL introduced the use of an electronic bulletin board where countries could post their scoring questions and receive scoring decisions from the domain experts. This information could be seen by all countries, who could then adjust their scoring.

To further ensure quality, countries were monitored as to the quality of their scoring in two ways.

First, within a country, at least 20 percent of the tasks had to be rescored. Guidelines for intra-country rescoring involved rescoring a larger portion of booklets at the beginning of the scoring process to identify and rectify as many scoring problems as possible. In a second phase, countries selected a smaller portion of the next third of the scoring booklets; this phase was viewed as a quality monitoring measure and involved rescoring a smaller portion of booklets regularly to the end of the rescoring activities. The two sets of scores needed to match with at least 95 percent accuracy before the next step of processing could begin. In fact, most of the intra-country scoring reliabilities were above 95 percent. Where errors occurred, a country was required to go back to the booklets and rescore all the questions with problems and all the tasks that belonged to a problem scorer.

Second, an international rescore was performed. Each country had 10 percent of its sample rescored by scorers in another country. For example, a sample of task booklets from the United States was rescored by the persons who had scored Canadian English booklets, and vice versa. The main goal of the rescore was to verify that no country scored consistently differently from another country. Intercountry score reliabilities were calculated by Statistics Canada and the results were evaluated by the ETS. Again, strict accuracy was demanded: a 90 percent correspondence was required before the scores were deemed acceptable. Any problems detected had to be rescored.

Estimation Methods

Weighting was used in ALL to adjust for sampling and nonresponse. Responses to the literacy tasks were scored using item response theory (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. Each participating country in ALL used a multistage probability sample design with stratification and unequal probabilities of respondent selection. Furthermore, there was a need to compensate for the nonresponse that occurred at varying levels. Therefore, the estimation of population parameters and the associated standard errors was dependent on the survey weights. All participating countries used the same general procedure for calculating the survey weights. However, each country developed the survey weights according to its

particular probability sample design. In general, two types of weights were calculated by each country: population weights that are required for the production of population estimates and jackknife replicate weights that are used to derive the corresponding standard errors.

Population weights. For each respondent record, the population weight was created first by calculating the theoretical or sample design weight, then by deriving a base sample weight by mathematically adjusting the theoretical weight for nonresponse. The base weight is the fundamental weight that can be used to produce population estimates. However, in order to ensure that the sample weights were consistent with a country's known population totals (i.e., benchmark totals) for key characteristics, the base sample weights were ratio-adjusted to the benchmark totals.

Jackknife weights. It was recommended that 10 to 30 jackknife replicate weights be developed for use in determining the standard errors of the survey estimates. Switzerland produced 15 jackknife replicate weights. The remaining countries produced 30 jackknife replicate weights.

Scaling. The results of ALL are reported along four scales—two literacy scales (prose and document), a single numeracy scale, and a scale capturing problem solving—with each ranging from 0 to 500 points. One might imagine these tasks arranged along their respective scale in terms of their difficulty for adults and the level of proficiency needed to respond correctly to each task. The procedure used in ALL to model these continua of difficulty and ability is IRT. IRT is a mathematical model used for estimating the probability that a particular person will respond correctly to a given task from a specified pool of tasks.

The scale value assigned to each item results from how representative samples of adults in participating countries perform on each item and is based on the theory that someone at a given point on the scale is equally proficient in all tasks at that point on the scale. For ALL, as for IALS, proficiency was determined to mean that someone at a particular point on the proficiency scale would have an 80 percent chance of answering items at that point correctly.

Just as adults within each participating country in ALL are sampled from the population of adults living in households, each task that was constructed and used in the assessment represents a type of task sampled from the domain or construct defined here. Hence, it is representative of a particular type of

literacy, numeracy, or problem-solving task that is associated with adult contexts.

In an attempt to display the progression of complexity and difficulty from the lower end of each scale to the upper end, each proficiency scale was divided into levels. Both the literacy and numeracy scales used five levels, where Level 1 represents the lowest level of proficiency and Level 5 the highest. These levels are defined as follows: 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). The scale for problem solving used four levels, where Level 1 is the lowest level of proficiency and Level 4 the highest. These four levels are defined as follows: Level 1 (0 to 250), Level 2 (251 to 300), Level 3 (301 to 350), and Level 4 (351 to 500).

Since each level represents a progression of knowledge and skills, individuals within a particular level not only demonstrate the knowledge and skills associated with that level but the proficiencies associated with the lower levels as well. In practical terms, this means that individuals performing at 250 (the middle of Level 2 on one of the literacy or numeracy scales) are expected to be able to perform the average Level 1 and Level 2 tasks with a high degree of proficiency. A comparable point on the problem-solving scale would be 275. In ALL, as in IALS, a high degree of proficiency is defined in terms of a response probability of 80 percent. This means that individuals estimated to have a particular scale score are expected to perform tasks at that point on the scale correctly with an 80 percent probability. It also means they will have a greater than 80 percent chance of performing tasks that are lower on the scale. It does not mean, however, that individuals with given proficiencies can never succeed at tasks with higher difficulty values. It does suggest that the more difficult the task relative to their proficiency, the lower the likelihood of a correct response.

Imputation. A respondent had to complete the background questionnaire, correctly complete at least two out of six simple tasks from 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, ALL countries were instructed to obtain at least a background questionnaire from

sampled individuals. ALL 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 ALL. 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). The reasons given most often by individuals for not completing the assessment were 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, ALL 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.

Future Plans

The OECD plans to conduct another survey, the Program for the International Assessment for Adult Competencies (PIAAC). It is built on the knowledge and experiences gained from IALS and ALL. PIAAC will measure relationships between educational background, workplace experiences and skills, professional attainment, use of ICT, and cognitive skills in the areas of literacy, numeracy and problem-solving. The assessment will be administered to 5,000 adults from ages 16 to 65. Administration of the survey will occur in 2011, with results being released in early 2013.

5. DATA QUALITY AND COMPARABILITY

The literacy tasks contained in ALL 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. ALL implemented procedures to minimize both sampling and nonsampling errors. The ALL sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. Quality control activities were employed during interviewer training, data collection, and processing of the survey data.

Sampling Error

Because ALL employed probability sampling, the results were subject to sampling error. Although small, this error was higher in ALL 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.

Nonsampling Error

The key sources of nonsampling error in ALL 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 ALL data

Table 17. Sample size and response rate for the United States for the Adult Literacy and Lifeskills Survey (ALL): 2003

Country	Population ages 16 to 65 (millions)	Initial sample size	Out-of- scope cases ¹	Number of respondents ²	Unweighted response rate (percent)
United States	184	7,045	1,846	3,420	66

¹Out-of-scope cases are those where the residents were not eligible for the survey, the dwelling could not be located, the dwelling was under construction, the dwelling was vacant or seasonal, or the cases were duplicates.

²A respondent's data are considered complete for the purposes of the scaling of a country's psychometric assessment data provided that at least the Background Questionnaire variables for age, gender, and education have been completed.

SOURCE: Desjardins, R., Murray, S., Clermont, Y., and Werquin, P. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey*. Ottawa, Canada: Statistics Canada.

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.

Coverage error. The design specifications for ALL 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 institutionalized 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 ALL countries attained a high level of population coverage. 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).

Nonresponse error. For ALL, 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 ALL 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 82 percent (Bermuda) to a low of 40 percent (Switzerland). The United States had an unweighted response rate of 66 percent (see [Table 17](#)).

Several precautions were taken against nonresponse bias. Interviewers were specifically instructed to return several times to nonrespondent households in order to obtain as many responses as possible. In

addition, all countries were asked to ensure that the address information provided to interviewers was as complete as possible in order to reduce potential household identification problems.

Item nonresponse. 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, numeracy, and problem solving) was well covered in terms of difficulty, stimuli type, and content domain. The ALL 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 final test design were presented to and approved by participating countries.

6. CONTACT INFORMATION

For content information on ALL, 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

General

Desjardins, R., Murray, S., Clermont, Y., and Werquin, P. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey*. Ottawa, Canada: Statistics Canada.

Lemke, M., Miller, D., Johnston, J., Krenzke, T., Alvarez-Rojas, L., Kastberg, D., and Jocelyn, L. (2005). *Highlights From the 2003 International Adult Literacy and Lifeskills Survey (ALL)- (Revised)* (NCES 2005-117rev). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.