

# National Adult Literacy Survey (NALS)

Website: <http://nces.ed.gov/naal/index.asp>

## 1. OVERVIEW

The National Adult Literacy Survey (NALS) was initiated to fill the need for accurate and detailed information on the English literacy skills of America's adults. In accordance with a congressional mandate, it provided the most detailed portrait that has ever been available in the 1990s on the condition of literacy in this nation.

The 1992 NALS is the third assessment of adult literacy funded by the federal government and conducted by the Educational Testing Service (ETS). The two previous efforts were (1) the 1985 Young Adult Literacy Assessment, funded as an adjunct to the National Assessment of Educational Progress (NAEP)—See [NAEP chapter](#); and (2) the Department of Labor's 1990 Workplace Literacy Survey. Building on these two earlier surveys, literacy for NALS is defined along three dimensions—prose, document, and quantitative—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 NALS provide a context for understanding the ways in which various characteristics are associated with demonstrated literacy skills.

NALS is the first national study of literacy for *all* adults since the Adult Performance Level Surveys conducted in the early 1970s. It is also the first in-person literacy assessment involving the prison population. A second adult literacy survey, the National Assessment of Adult Literacy (NAAL), was conducted in 2003.

### Purpose

To (1) evaluate the English language literacy skills of adults (16 years and older) living in households or prisons in the United States; (2) relate the literacy skills of the nation's adults to a variety of demographic characteristics and explanatory variables; and (3) compare the results with those from the 1985 Young Adult Literacy Assessment and the 1990 Workplace Literacy Survey.

### Components

The 1992 survey consisted of one component that was administered to three different representative samples: a national household sample; supplemental state household samples for 12 states (California, Florida, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, and Washington); and a national sample of federal and state prison inmates. Responses from the national, state, and prison samples were combined to yield the best possible performance estimates.

**National Adult Literacy Survey.** The 1992 survey assessed the literacy skills of a representative sample of the U.S. adult population using simulations of three kinds of literacy tasks that adults would ordinarily encounter in daily life (prose, document, and quantitative literacy). The data were collected through in-person interviews with adults who were living in households or in federal or state prisons. Adults were defined as individuals 16 years or older for the national and prison samples, and 16 to 64 years of age for the state samples. In addition to the cognitive

### PERIODIC SURVEY OF A SAMPLE OF ADULTS LIVING IN HOUSEHOLDS OR PRISONS

Assesses literacy skills:

- Prose
- Document
- Quantitative

Collects background data:

- Demographics
- Education
- Labor market experiences
- Income
- Activities

tasks, the personal interview gathered information on demographic characteristics, language background, educational background, reading practices, and labor market experiences. To ensure comparability across all samples, the literacy tasks assessed were the same for all three samples. Background data varied somewhat between the household and prison samples—labor force questions were irrelevant to prisoners, and questions about criminal behavior and sentences were relevant only to prisoners.

**Literacy Assessment.** The pool of literacy tasks used to measure adult proficiencies consisted of 165 literacy questions—41 prose, 81 document, and 43 quantitative. To ensure that valid comparisons could be made by linking the scales to those of the 1985 Young Adult Literacy Assessment, 85 tasks from that survey were included in the 1992 survey. An additional 80 new tasks were developed specifically to complement and enhance the original 85 tasks. The literacy tasks administered in NALS varied widely in terms of materials and content. The six major context/content areas were home and family; health and safety; community and citizenship; consumer electronics; work; and leisure and recreation. Each adult was given a subset (about 45) of the total pool of assessment tasks to complete. Each of the tasks extended over a range of difficulty on the three literacy scales. The new tasks were designed to simulate the way in which people use various types of materials and to require different strategies for successful performance.

The responses to the literacy assessment were pooled and reported by proficiency scores, ranging from 0 to 500, on three separate scales, one each for prose, document, and quantitative literacy. By examining the overall characteristics of individuals who performed at each literacy level on each scale, it is possible to identify factors associated with higher or lower proficiency in reading and using prose, document, and quantitative materials.

**Background Information.** Background information collected for the state and household samples included data on *background and demographics*—country of birth, languages spoken or read, access to reading materials, size of household, educational attainment of parents, age, race/ethnicity, and marital status; *education*—highest grade completed in school, current aspirations, participation in adult education classes, and education received outside the country; *labor market experiences*—employment status, recent labor market experiences, and occupation; *income*—personal and household; and *activities*—voting behavior, hours spent watching television, frequency and content of newspaper reading, and use of literacy skills for work

and leisure. Respondents from each of the 12 participating states were also asked state-specific questions.

To address issues of particular relevance to the prison population, a separate background questionnaire was developed for the prison sample. This instrument drew questions from the 1991 Survey of Inmates of State Correctional Facilities, sponsored by the Department of Justice's Bureau of Justice Statistics. The background questionnaire for the prison population addressed the following major topics: general and language background; educational background and experience; current offenses and criminal history; prison work assignments and labor force participation prior to incarceration; literacy activities and collaboration; and demographic information.

### Periodicity

NALS was conducted in 1992. NAAL, a continuation of NALS, was conducted in 2003.

## 2. USES OF DATA

Results from NALS provide a detailed portrait on the condition of literacy in this nation. NALS data provide vital information to policymakers, business and labor leaders, researchers, and citizens. The survey results can be used to

- describe the levels of literacy demonstrated by the adult population as a whole and by adults in various subgroups (e.g., those targeted as at risk, prison inmates, and older adults);
- characterize adults' literacy skills in terms of demographic and background information (e.g., reading characteristics, education, and employment experiences);
- profile the literacy skills of the nation's workforce;
- compare assessment results from the current study with those from the 1985 Young Adult Literacy Assessment;
- interpret the findings in light of information-processing skills and strategies, so as to inform curriculum decisions concerning adult education and training; and
- increase understanding of the skills and knowledge associated with living in a technological society.

### 3. KEY CONCEPTS

Some of the key concepts related to the literacy assessment are described below. See the NALS Electronic Codebook or appendices of NALS reports for lists and descriptions of variables.

**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. This definition goes beyond simply decoding and comprehending text to include a broad range of information-processing skills that adults use in accomplishing the range of tasks associated with work, home, and community contexts.

**Prose Literacy.** The ability to locate information contained in expository or narrative prose in the presence of related but unnecessary information, find all of the relevant information, integrate information from various parts of a passage of text, and write new information related to the text. Expository prose consists of printed information in the form of connected sentences and longer passages that define, describe, or inform, such as newspaper stories or written instructions. Narrative prose tells a story, but is less frequently used by adults in everyday life than by school children, and did not occur as often in the text presented in NALS as prose literacy tasks. Prose varies in its length, density, and structure.

**Document Literacy.** The ability to locate information in documents, repeat the search as many times as needed to find all the information, integrate information from various parts of a document, and write new information as requested in appropriate places in a document, while screening out related but inappropriate information. Documents differ from prose text in that they are more highly structured. Documents consist of structured prose and quantitative information in complex arrays arranged in rows and columns, such as tables, data forms, and lists (simple, nested, intersected, or combined); in hierarchical structures, such as tables of contents or indexes; or in two-dimensional visual displays of quantitative information, such as graphs, charts, and maps.

**Quantitative Literacy.** The ability to use quantitative information contained in prose or documents (specifically the ability to locate quantities while screening out related but unneeded information), repeat the search as many times as needed to find all the numbers, integrate information from various parts of a text or document, infer the necessary arithmetic operation(s), and perform arithmetic operation(s). Quantities can be located in either prose texts or in

documents. Quantitative information may be displayed visually in graphs, maps, or charts, or it may be displayed numerically using whole numbers, fractions, decimals, percentages, or time units (hours and minutes).

**Literacy Scales.** 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 1985 Young Adult Literacy Assessment. The scores on each scale represent degrees of proficiency along that particular dimension of literacy. The literacy tasks administered in the 1992 survey 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 in English 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.

**Succeed Consistently.** Indicates that a person at or above a given level of literacy has at least an 80 percent chance of correctly responding to a particular task. This 80 percent criterion is more stringent than the 65 percent standard used in NAEP (see [NAEP chapter](#)) for measuring what school children know and can do.

### 4. SURVEY DESIGN

The 1992 NALS was designed and administered by ETS. A subcontract was awarded to Westat, Inc., for sampling and field data collection. A committee of experts from business and industry, labor, government, research, and adult education worked with the ETS staff to develop the definition of literacy that underlies NALS, as well as to prepare the assessment objectives that guided the selection and construction of assessment tasks. In addition to this Literacy Definition Committee, a Technical Review Committee was formed to help ensure the soundness of the assessment design, the quality of the data collected, the integrity of

the analyses conducted, and the appropriateness of the interpretations of the final results. The prison survey was developed in consultation with the Bureau of Justice Statistics and the Federal Bureau of Prisons. The survey design for the 1992 survey is described below.

### Target Population

The target population for the national household sample consisted of adults 16 years and older in the 50 states and the District of Columbia who, at the time of the survey, resided in private households or college dormitories. The target population for the supplemental state household sample consisted of individuals 16 to 64 years of age who, at the time of the survey, resided in private households or college dormitories in the participating state (California, Florida, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, or Washington). Individuals residing in other institutions—nursing homes, group homes, or psychiatric facilities—were not included in the household samples. The target population for the prison sample consisted of adults 16 years or older who were in state or federal prisons at the time of the survey; those held in local jails, community-based facilities, or other types of institutions were not included.

### Sample Design

Because this 1992 survey was designed to provide data representative at the national level (including prison inmates) and at the state level for participating states, it included three different samples: a national household sample, supplemental state household samples for 12 states, and a supplemental national sample of state and federal prison inmates.

**Household Samples.** The sample design for the national and state household samples involved a four-stage stratified area sample: (1) the selection of primary sampling units (PSUs) consisting of counties or contiguous groups of counties; (2) the selection of segments (within the selected PSUs) consisting of census blocks or groups of contiguous census blocks; (3) the selection of households within the segmented samples; and (4) the selection of age-eligible individuals within each selected household. The sample design requirements called for an average cluster size of seven interviews (i.e., seven completed background questionnaires per segment). In addition, a reserve sample at the household level of approximately 5 percent of the size of the main sample was selected and set aside in case of shortfalls due to unexpectedly high vacancy and nonresponse rates.

One national area sample was drawn for the national household sample, and 12 independent state-specific area samples were drawn from the 12 states participating in the supplemental state samples. The sample designs used for all 13 samples were similar, with one major difference. In the national sample, Black and Hispanic respondents were sampled at about double the rate of the remainder of the population to assure reliable estimates of their literacy proficiencies, whereas the state samples used no oversampling.

The first stage of sampling involved the selection of PSUs. A national sampling frame of 1,400 PSUs was constructed primarily from 1990 census data stratified on the basis of region, metropolitan status, percent Black, percent Hispanic, and whenever possible, per capita income. Using this frame, 101 PSUs were selected for the national sample. The national frame of PSUs (subdivided at state boundaries, if needed) was used to construct individual state frames for the supplemental state sample; a sample of 8 to 12 PSUs was selected within each of the given states. All PSUs were selected with probability proportional to the PSU's 1990 population.

The second stage of sampling involved the selection of segments within the selected PSUs. The Bureau of the Census's Topologically Integrated Geographical Encoding and Referencing (TIGER) System File was used for the production of segment maps. The segments were selected with probability proportional to size, where the measure of size for a segment was a function of the number of year-round housing units within the segment. The oversampling of Black and Hispanic respondents for the national sample was carried out at the segment level, where segments were classified either as having a high percentage of the Black or Hispanic population (more than 25 percent) or as not having a high percentage.

The third stage of sampling involved the selection of households within the segmented samples. Westat field staff visited all selected segments in the fall of 1991 and prepared lists of all housing units within the boundaries of each segment as determined by the 1990 census block maps. The lists were used to construct the sampling frame for households. Households were selected with equal probability within each segment, except for White, non-Hispanic households in segments with a high percentage of the Black or Hispanic population (over 25 percent) in the national sample, which were subsampled so that the sampling rates for White, non-Hispanic respondents would be about the same overall.

The fourth stage of sampling involved the selection of one or two adults within each selected household during the data collection phase of the survey. One person was selected at random from households with fewer than four eligible members; two persons were selected from households with four or more eligible members. Using a screener, the interviewer constructed a list of age-eligible household members (16 and older for the national sample, 16 to 64 for the state sample) for each selected household. The interviewers, who were instructed to list the eligible household members in descending order by age, then identified one or two household members to interview, based on computer-generated sampling messages that were attached to each questionnaire in advance.

**Prison Sample.** There were two stages of selection for the prison sample. The first stage involved the selection of state or federal correctional facilities. The sampling frame for the correctional facilities was based on the 1990 census of federal and state prisons, updated in mid-1991. The facility frame was stratified prior to sample selection on the basis of type of facility (federal or state prison), region of country, inmate gender composition, and type of security. A sample of 88 facilities and a reserve sample of 8 facilities was then drawn from the frame based on probability proportional to size, where the measure of size for a given facility was equal to the inmate population. The second stage of sampling involved the selection of inmates within each selected facility, using a list of names obtained from the facility administrators. An average of 12 inmates were selected from each facility based on a probability inversely proportional to their facility's inmate population (up to a maximum of 22 interviews in a facility), so that the product of the first- and second-stage probabilities would be constant.

### Assessment Design

Building on the 1985 Young Adult Literacy Assessment and the 1991 Workplace Literacy Survey, the NALS Technical Committee adopted the definition of literacy and the literacy scales—prose, document, and quantitative—used in the previous surveys. The materials were selected to represent a variety of contexts and contents: home and family; health and safety; community and citizenship; consumer electronics; work; and leisure and recreation.

**BIB Spiraling.** The survey design gave each respondent a subset of the total pool of literacy tasks, while at the same time ensuring that each of the 165 tasks was administered to a nationally representative sample of the adult population. The design most suitable for this purpose is a variant of standard matrix

sampling called balanced incomplete block (BIB) design.

Literacy tasks were assigned to blocks or sections that could be completed in about 15 minutes, and these blocks were then compiled into booklets so that each block appeared in each position (first, middle, and last) and each block was paired with every other block. Thirteen blocks of simulation tasks were assembled into 26 unique booklets, each of which contained four blocks of tasks: the core (the same for all exercise booklets) and three cognitive blocks. Each booklet could be completed in about 45 minutes.

**Pretests.** A field test of the national household sample was conducted in the spring of 1991 using a sample of 2,000 adults drawn from 16 PSUs. The purposes of the field test were to evaluate the impact of incentives on response rates, performance, and survey costs; to evaluate newly developed literacy exercises for item bias and testing time; and to evaluate the administration and appropriateness of the background questions. As a result of the field test, some of the literacy tasks and their scoring guides were revised or dropped from the final assessment.

For the prison sample, a small pretest was conducted at the Roxbury Correctional Institution in Hagerstown, Maryland. This pretest was designed to evaluate the ease of administration of the survey instruments, survey administration time, within-facility procedures, and inmate reaction to the survey. The pretest demonstrated that several changes to the background questionnaire would facilitate administration. Administrative procedures were also refined to reflect lessons learned during the pretest.

### Data Collection and Processing

The survey data were collected through in-person household or prison interviews during the first 8 months of 1992. As field operations were completed, the data were shipped to ETS for processing. Further description follows.

**Reference Dates.** Respondents answered the employment status and weekly wages questions for the week before the survey was administered.

**Data Collection.** During January and February of 1992, field interviewers, supervisors, and editors received extensive training both in general and survey-specific interview techniques. The NALS field period began in February 1992, immediately following the completion of the first interviewer training sessions, and lasted 28 weeks, until the end of August. All three survey sample groups were worked simultaneously (except for the state of Florida, where data were not collected until

1993). Except for a small, experimental “no incentive” group, all household participants who completed as much of the assessment as their skills allowed received \$20 for their time. More than 400 trained interviewers visited about 44,000 households to select and interview almost 31,000 adults. In addition, over 1,147 prison inmates at 87 facilities were interviewed.

Each survey participant was asked to spend approximately one hour responding to survey questions and tasks. Data collection instruments included the screener (designed to enumerate household members and select survey respondents), the background questionnaire, and the literacy exercise booklets. Answering the screener and background questionnaire required no reading or writing skills; to ensure standardized administration, the questions on each were read to respondents in English or Spanish and the answers recorded by the assessment interviewer. Each of the exercise booklets had a corresponding interview guide, with specific instructions to the interviewer for directing the exercise booklet. Reading and writing skills in the English language were required to complete the exercise booklet. When a sampled respondent did not complete any or all of the survey instruments, the interviewer was required to complete a noninterview report form. Field supervisors reviewed the noninterview forms to determine the case’s potential for conversion, and the data collected on the form were processed for nonresponse analysis.

Following the completion of an interview, interviewers edited all materials for legibility and completeness. The interviewers sent their completed work to their regional supervisors for a complete edit of the instruments, quality control procedures, and any required data retrieval. As these tasks were completed, the cases were shipped to ETS for processing.

During the data collection process, two special quality control procedures were implemented to identify any households or dwellings missed during the listing phase: the missing structure procedure and the missed dwelling unit procedure. These procedures were used to give these missed structures and dwelling units a chance of selection at time of data collection.

The field effort occurred in three overlapping stages:

- (1) *Initial Phase.* Each area segment was assigned by the regional supervisor to an interviewer, who followed certain rules in making a prescribed number of calls (a maximum of four was used) to every sampled dwelling in the segment.
- (2) *Reassignment Phase.* Cases that did not result in completed interviews during the initial phase

were reviewed by the regional supervisor, and a subset was selected for reassignment to another interviewer in the same PSU or an interviewer from a nearby PSU.

- (3) *Special Nonresponse Conversion Phase.* The home office assembled a special traveling team of the most experienced or productive interviewers to perform a nonresponse conversion effort, under the supervision of a subset of the field supervisors.

**Data Processing.** Coding and scoring staff underwent intensive training prior to the actual coding and scoring. A scoring supervisor monitored both the coding of the questionnaires and the scoring of the exercise booklets. The background questionnaire was designed to be read by a computerized scanning device. Nearly all the simulation tasks contained in the exercise booklet were open-ended; with scoring guides as examples, responses to these items were classified as correct, incorrect, or omitted by trained readers. Responses from the screener and scores from the exercise booklets were transferred to scannable answer sheets. Each survey instrument’s scannable forms were batched and sent to the scanning department at regular intervals. As the different instruments were processed, the data were transferred to a database on the main ETS computer for editing.

**Editing.** Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer’s closeout work, and field observation of both supervisors and interviewers. Additional edits were done during data processing. These included an assessment of the internal logic and consistency of the data received. Discrepancies were corrected whenever possible. The background questionnaires were also checked to make sure that the skip patterns had been followed and all data errors were resolved. In addition, a random set of exercise booklets was selected to provide an additional check on the accuracy of transferring information from booklets and answer sheets to the database.

### **Estimation Methods**

Weighting was used in the 1992 NALS, prior to the calculation of base weights. 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. An innovative approach was implemented to impute missing cognitive data in order

to minimize distortions in the population proficiency estimates due to nonresponse to the literacy booklet.

**Weighting.** Full sample and replicate weights were calculated for survey respondents who completed the exercise booklet; those who could not start the exercises because of a language barrier, a physical or mental barrier, or a reading or writing barrier; and those who refused to complete the exercises but had completed background questionnaires. Demographic variables critical to the weighting were recoded and imputed, if necessary, prior to the calculation of base weights (see “[Imputation](#)” below). Separate sets of weights were computed for the incentive and “no incentive” samples.

**Household samples.** A base weight was computed for each eligible record. The base weight initially was computed as the reciprocal of the product of probabilities of selection for a respondent at the PSU, segment, dwelling unit, and person levels. The final base weight included adjustments to reflect the selection of the reserve sample, the selection of missed dwelling units, and the chunking process conducted during the listing of the segments; and to account for the subsample of segments assigned to the “no incentive” experiment and the subsampling of respondents within households. The base weights for each sample were then poststratified to known 1990 census population totals, adjusted for undercount. This first-level stratification provided sampling weights with lower variation and adjusted for nonresponse. State records were poststratified separately from national records to provide a common base for applying composite weighting factors; population totals were calculated separately for each distinct group.

Composite weights were developed so that NALS data could be used to produce both state and national statistics. For the household samples, a composite weight was computed as the product of the poststratified base weight and a compositing factor that combined the national and state sample data in an optimal manner, considering the differences in sample design, sample size, and sampling error between the two sampled groups. Up to four different compositing factors were used in each of the 11 participating states, and a pseudo-factor (equal to 1) was used for all persons 65 and older and for all national sample records from outside the 11 participating states.

To compute the final sample weights, the composite weights were adjusted to known 1990 census counts (adjusted for undercount), using a process called the poststratification raking ratio adjustment. The cells used for raking were defined to the finest combination

of age, race/ethnicity, sex, education, and geographic indicators (e.g., Metropolitan Statistical Area [MSA] vs. non-MSA) that the data would allow. Raking adjustment factors were calculated separately for each of the state samples and then for the remainder of the United States.

The above steps used to create the final sample weights were repeated for 60 strategically constructed subsets of the household sample to create a set of replicate weights to be used for variance estimation using the jackknife method.

**Prison sample.** Base weights for the prison respondents were constructed to be equal to the reciprocal of the product of the selection probabilities for the facility and the inmate within the facility. These weights were then nonresponse-adjusted to reflect both facility and inmate nonresponse. To compute the final sample weights, the resulting nonresponse-adjusted weights were then raked to agree with independent estimates for certain subgroups of the prison population. The above procedures were repeated for 45 strategically constructed subsets of the prison sample to create a set of replicate weights to be used for variance estimation using the jackknife method.

**Scaling.** Since NALS used a variant of matrix sampling and since different respondents received different sets of tasks, it would be inappropriate to report its results using conventional scoring methods based on the number of correct responses. The literacy assessment results are reported using IRT scaling, which assumes some uniformity in response patterns when items require similar skills. Such uniformity can be used to characterize both examinees and items in terms of a common scale attached to the skills, even when all examinees do not take identical sets of items. Comparisons of items and examinees can then be made in reference to a scale, rather than to the percent correct. IRT scaling also allows the distributions of examinee groups to be compared.

The results of the 1992 literacy assessment are reported on three scales (prose, document, and quantitative) that were established for the 1985 Young Adult Literacy Assessment. Separate IRT linking and scaling were carried out for each of the three domains, using the three-parameter logistic (3PL) scaling model from item response theory. This is a mathematical model for estimating the probability that a particular person will respond correctly to a particular item from a single domain of items. The probability is given as a function of a parameter characterizing the proficiency of that person and three parameters characterizing the properties of that item. Item parameters needed for the

3PL scaling model were estimated by linking each of the literacy scales used in the 1992 survey to the 1985 Young Adult Literacy Assessment scales.

**Imputation.** Imputation was performed prior to weighting on missing demographic items considered critical to weighting. Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Missing cognitive data were also imputed.

**Demographic data.** Demographic variables critical to the weighting (race/ethnicity of the head of household; sex, age, race/ethnicity, and education of the respondent) were recoded and collapsed to required levels, and imputed, if necessary, prior to the calculation of base weights. Data from the background questionnaire were preferred for all items except race/ethnicity of the head of household, which was collected in the screener. For the few cases in which the background questionnaire measure was missing, the screener measure was generally available and was used as a direct substitute. The amount of missing data remaining after substitution was small, making the imputation task fairly straightforward. A standard (random within class) hot-deck imputation procedure was performed for particular combinations of fields that were missing. Imputation flags were created for each of the five critical fields to indicate whether data were originally reported or were based on substitution or imputation. The imputed values were used only for the sample weighting process.

**Literacy proficiency estimation (plausible values).** A multiple imputation procedure based on plausible values methodology was used to estimate respondents' literacy proficiency in the 1992 NALS. When analyzing the distribution of proficiencies in a group of persons, more efficient estimates can be obtained from a sample design similar to that used in this 1992 survey. Such designs solicit relatively few cognitive responses from each sampled respondent, but maintain a wide range of content representation when responses are summed for all respondents.

In the 1992 survey, all proficiency data were based on two types of information: responses to the background questions and responses to the cognitive items. As an intermediate step, a functional relationship between the two sets of information was calculated for the total sample, and this function was used to obtain unbiased proficiency estimates for population groups with reduced error variance. Possible values for a respondent's proficiency were sampled from a posterior distribution that is the product of two

functions: the conditional distribution of proficiency given the pattern of background variables and the likelihood function of proficiency given the pattern of responses to the cognitive items. Since exact matches of background responses are quite rare, NALS used more than 200 principal components to summarize the background information, capturing more than 99 percent of the variance. More detailed information on the plausible values methodology used in the 1992 survey is available in the [Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey](#) (Kirsch et al. 2000).

**Cognitive data.** New procedures were implemented in the 1992 NALS to minimize distortions in the population proficiency estimates due to nonresponse to the literacy booklets. When a sampled individual decided to stop the assessment (answered less than five literacy items per scale), 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 half 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.

To represent the range of implied causes of missing literacy responses, the imputation procedure selected relied on background variables and self-reported reasons for nonresponse, in addition to the functional relationship between background variables and proficiency scores for the total population. It treated "consecutively missing" data from the literacy booklet instrument differently depending on whether the nonrespondents' reasons were related or unrelated to their literacy skills: (1) those who gave literacy-related reasons were treated as wrong answers, based on the assumption that they could not have correctly completed the literacy tasks, whereas (2) those who gave no reason or cited reasons unrelated to literacy skills for not completing the assessment were essentially ignored (considered not reached), since it could not be assumed that their answers would have been either correct or incorrect. The proficiencies of such respondents were inferred from the proficiencies of other adults with similar characteristics using the plausible values methodology described above.

### Future Plans

A second survey, NAAL, was conducted in 2003. Currently, there are no plans to administer another measure of adult literacy.

## 5. DATA QUALITY AND COMPARABILITY

The NALS sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. In addition, NCES conducted special evaluation studies to examine issues related to the quality of NALS. These studies included (1) a study of the role of incentives in literacy survey research; (2) an evaluation of its sample design and composite estimation; and (3) an evaluation of the construct validity of the adult literacy scales.

### Sampling Error

In the 1992 survey, the use of a complex sample design, adjustments for nonresponse, and poststratification procedures resulted in dependence among the observations. Therefore, a jackknife replication method was used to estimate the sampling variance. The mean square error of replicate estimates around their corresponding full sample estimate provides an estimate of the sampling variance of the statistic of interest. The replication scheme was designed to produce stable estimates of standard errors for national and prison estimates as well as for the 12 individual states.

The advantage of compositing the national and state samples during sample weighting was the increased sample size, which improved the precision of both the state and national estimates. However, biases could be present because the national PSU sample strata were not designed to maximize the efficiency of state-level estimates.

### Nonsampling Error

The major source of nonsampling error in the 1992 NALS was nonresponse error; special procedures were developed to minimize potential nonresponse bias based on how much of the survey the respondent completed. Other possible sources of nonsampling error were random measurement error and systematic error due to interviewers, coders, or scorers.

**Coverage Error.** Coverage error could result from either the sampling frame of households or prisons being incomplete or from a household's or prison's failure to include all adults 16 years and older on the lists from which the sampled respondents were drawn. Special procedures and edits were built into NALS to

review both listers' and interviewers' ongoing work and to give any missed structures and/or dwelling units a chance of selection at data collection. However, just as all other household personal interview surveys have persistent undercoverage problems, the 1992 survey had problems in population coverage due to interviewers not gaining access to households in dangerous neighborhoods, locked residential apartment buildings, and gated communities.

### *Nonresponse Error.*

*Unit nonresponse.* Since three survey instruments— screener, background questionnaire, and exercise booklet—were required for the administration of the survey, it was possible for a household or respondent to refuse to participate at the time of the administration of any one of these instruments. Because the screener and background questionnaire were read to the survey participants in English or Spanish, but the exercise booklet required reading and writing in the English language, it was possible to complete the screener or background questionnaire but not the exercise booklet, and vice versa. Thus, response rates were calculated for each of the three instruments for the household samples (see table 12). For the prison sample, there were only two points at which a respondent could not respond—at the administration of the background questionnaire or the exercise booklet.

The response rate to the background questionnaire was 80.5 percent. For the household samples, the response rates exclude individuals who were not paid incentives. Also excluded are the respondents to the Florida state survey, which had a delayed administration.

The combined national and state household target sample in the 1992 NALS included 43,780 representative housing units, of which 5,410 were vacant. Approximately 89 percent of the occupied households completed a screener.

The household sample screening effort identified a total of 30,810 eligible respondents, of whom 24,940 (81.0 percent unweighted) completed the background questionnaire. For the prison sample, 87 of the 88 sampled facilities participated in the survey. Of the 1,340 inmates selected, 1,150 (85.6 percent unweighted) completed the background questionnaire.

For the occupied households, "refusal or breakoff" was the most common explanation for nonresponse to the screener and background questionnaire. The second most common explanation was "not at home after maximum number of calls." Nonresponse also resulted from language, physical, and mental problems. Housing units or individuals who refused to participate before any information was collected about them, or

who did not answer a sufficient number of background questions, were never incorporated into the database. Because these individuals were unlikely to know that the survey intended to assess their literacy, it was assumed that their reason for not completing the survey was not related to their level of literacy.

Literacy assessment booklets were considered complete if at least five items were answered on each scale. A total of 24,940 household sample members were classified as eligible for the exercise booklet. Of these, 88.6 percent completed the booklet and another 6.1 percent partially completed it. Of the 1,150 eligibles in the prison sample, 86.8 percent completed the booklet and another 9.3 percent partially completed it.

There were reasons to believe that the literacy performance data were missing more often for adults with lower levels of literacy than for adults with higher levels. Field-test evidence and experience with surveys indicated that adults with lower levels of literacy were more likely than adults with higher proficiencies either to decline to respond to the survey at all or to begin the assessment but not complete it. Ignoring this pattern of missing data would have resulted in overestimating the literacy skills of adults in the United States. Therefore, to minimize bias in the proficiency estimates due to nonresponse to the literacy assessment, special procedures were developed to impute the literacy proficiencies of nonrespondents who completed fewer than five literacy tasks.

*Item nonresponse.* For each background questionnaire, staff verified that certain questions providing critical information for weighting and data analyses had been answered, namely, education level, employment status, parents' level of education, race, and sex. If a response was missing, the case was returned to the field for data retrieval. Therefore, item response rates for completed background questionnaires were quite high, although they varied by type of question. Questions asking country of origin (first question in the booklet) and sex (last question in the booklet) had nearly 100 percent response rates, indicating that most respondents attempted to complete the entire questionnaire. Response rates were lower, however, for questions about income and educational background.

The electronic codebook provides counts of item nonresponse. These, however, have to be considered in terms of the number of adults that were offered each task, because a great deal of the missing data is missing by design.

**Measurement Error.** All background questions and literacy tasks underwent extensive review by subject

**Table 12. Weighted and unweighted response rates for all sample types in the National Adult Literacy Survey, by survey component: 1992**

Component	Weighted (percent)	Unweighted (percent)
Screener	—	89.1
Background questionnaire	80.5	81.0
Exercise booklet	95.9	95.9

— Not available.

NOTE: The weighted response rates were calculated by applying the sampling weight to each individual to account for his or her probability of selection into the sample. Weighted response rates were computed only for screened households (the probability of selection is not known for persons in households that were not screened).

SOURCE: Kirsch, I.S., Yamamoto, K., Norris, N., Rock, D., Jungeblut, A., O'Reilly, P., Campbell, A., Jenkins, L., Kolstad, A., Berlin, M., Mohadjer, L., Waksberg, J., Goksel, H., Burke, J., Rieger, S., Green, J., Klein, M., Mosenthal, P., and Baldi, S. (2000). *Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey* (NCES 2001-457). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

area and measurement specialists, as well as scrutiny to eliminate any bias or lack of sensitivity to particular groups. Special care was taken to include materials and tasks that were relevant to adults of widely varying ages. During the test development stage, the tasks were submitted to test specialists for review, part of which involved checking the accuracy and completeness of the scoring guide. After preliminary versions of the assessment instruments were developed and after the field test was conducted, the literacy tasks were closely analyzed for bias or "differential item functioning." The goal was to identify any assessment tasks that were likely to underestimate the proficiencies of a particular subpopulation, whether it be older adults, females, or Black or Hispanic adults. Any assessment item that appeared to be biased against a subgroup was excluded from the final survey. The coding and scoring guides also underwent further revisions after the first responses were received from the main data collection.

*Interviewer error checks.* Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer's closeout work, and field observation of both supervisors and interviewers.

*Coding/scoring error checks.* In order to monitor the accuracy of coding, the questions dealing with country of birth, language, wages, and date of birth were

checked in 10 percent of the questionnaires by a second coder. For the industry and occupation questions, 100 percent of the questionnaires were recoded by a second coder. Twenty percent of all the exercise booklets were subjected to a reader reliability check, which entailed a scoring by a second reader. There was a high degree of reader reliability across tasks—ranging from 88.1 to 99.9 percent—with an average agreement of 97 percent. For 133 out of 165 open-ended tasks, the agreement between the two readers was above 95 percent.

### Data Comparability

One of the major goals of this survey was to compare its results to the 1985 Young Adult Literacy Assessment and other large assessment studies. NALS is also comparable with NAAL, conducted in 2003, in terms of assessment scores (see [NAAL chapter](#)).

**Comparisons with the 1985 Young Adult Literacy Assessment.** Comparisons are possible because the sample design, item pool, and methodology used in the 1985 Young Adult Literacy Assessment and the 1992 survey were very similar. Literacy tasks for each survey were developed using the same definition of literacy, and a subset of identical tasks was administered in both assessments. Scoring guides were the same for both surveys. Both gave nearly identical incentive payments to participants (\$15 in 1985 and \$20 in 1992). The literacy scales used in the two surveys were linked so that the scores could be reported on a common scale.

Nevertheless, there were some differences in procedures for the two surveys. For example, missing responses to the literacy tasks were handled differently. In the 1985 Young Adult Literacy Assessment, individuals who could not answer six core literacy tasks and those who spoke only Spanish were excluded from the analyses. In the 1992 survey, however, a special procedure was used to impute literacy proficiencies for literacy-related nonrespondents.

Due to such procedural differences, direct comparisons of the results of the two surveys are not simple and straight-forward. However, because the 1992 sample is more inclusive than the 1985 sample, subsamples that have more exact counterparts in the 1985 survey can be selected. For instance, the initial report from the 1992 NALS presented data, using no subsample matching that indicated that young adults in 1992 were somewhat less literate than their predecessors in 1985. However, when a comparison was made between matched subsamples of the 1985 and 1992 survey respondents based on reasons for nonresponse, the proficiency differences decreased significantly.

Furthermore, results from partition analysis of the two surveys' matched subsamples—based on change due to variations in demographic characteristics versus change not related to demography—suggest that most of the observed declines in the average literacy skills of young adults over time can be accounted for by shifts in the composition of the population and by changes across the assessments in the rules used to include or exclude nonrespondents.

**Comparisons with the 1993 General Educational Development (GED) Tests.** Comparisons between NALS and GED examinees are explored in *The Literacy Proficiencies of GED Examinees: Results From the GED-NALS Comparison Study* (Baldwin et al. 1993). The GED tests and NALS instruments have a considerable degree of overlap in what they measure. Both assess skills that appear to represent verbal comprehension and reasoning or the ability to understand, analyze, interpret, and evaluate written information and apply fundamental principles and concepts. Despite the considerable degree of overlap, the two instruments also measure somewhat different skills. For example, the GED tests seem to tap unique dimensions of writing mechanics and mathematics, while the adult literacy scales appear to tap unique dimensions of document literacy. In addition, the evidence shows that there are no differences in the average prose, document, or quantitative literacy skills of those adults who terminated their schooling at the high school or GED level.

## 6. CONTACT INFORMATION

For content information on the National Adult Assessments of Literacy, contact

Andrew J. Kolstad  
Phone: (202) 502-7374  
E-mail: [andrew.kolstad@ed.gov](mailto:andrew.kolstad@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

Baldwin, J., Kirsch, I.S., Rock, D., and Yamamoto, K. (1993). *The Literacy Proficiencies of GED*

*Examinees: Results From the GED-NALS Comparison Study.* American Council on Education and Educational Testing Service.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, K. (1993). *Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey* (NCES 93-275). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Kirsch, I.S., Yamamoto, K., Norris, N., Rock, D., Jungeblut, A., O'Reilly, P., Campbell, A., Jenkins, L., Kolstad, A., Berlin, M., Mohadjer, L., Waksberg, J., Goksel, H., Burke, J., Rieger, S.,

Green, J., Klein, M., Mosenthal, P., and Baldi, S. (2000). *Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey* (NCES 2001-457). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

### **Survey Design**

Campbell, A., and Kirsch, I.S. (1992). *Assessing Literacy: The Framework for the National Adult Literacy Survey* (NCES 92-113). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.