

National Assessment of Adult Literacy (NAAL)

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

1. OVERVIEW

The 2003 National Assessment of Adult Literacy (NAAL) is a nationally representative assessment of English literacy among American adults age 16 and older. Sponsored by the National Center for Education Statistics (NCES), NAAL is the nation's most comprehensive measure of adult literacy since the 1992 National Adult Literacy Survey (NALS).

In 2003, over 19,000 adults participated in the national and state-level assessments, representing the entire population of U.S. adults age 16 and older (in households and prisons) in the 50 states and the District of Columbia. Approximately 1,200 of the participants were inmates of state and federal prisons who were assessed separately in order to provide estimates of literacy for the incarcerated population.

By comparing results from 1992 and 2003, NAAL provides the first indicator in a decade of the nation's progress in adult literacy. NAAL also provides information on adults' literacy performance and related background characteristics to researchers, practitioners, policymakers, and the general public.

Purpose

To (1) evaluate the English language literacy skills of adults (age 16 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 1992 NALS.

Components

NAAL includes a number of components that capture the breadth of adult literacy in the United States: the Background Questionnaire helps identify the relationships between adult literacy and selected demographic and background characteristics; the Prison Component assesses the literacy skills of adults in federal and state prisons; the State Assessment of Adult Literacy (SAAL) gives statewide estimates of literacy for states participating in the state-level assessment; the Health Literacy Component introduces the first-ever national assessment of adults' ability to use their literacy skills in understanding health-related materials and forms; the Fluency Addition to NAAL (FAN) measures basic reading skills by assessing adults' ability to decode, recognize words, and read with fluency; the Adult Literacy Supplemental Assessment (ALSA) provides information on the ability of the least literate adults to identify letters and numbers and to comprehend simple prose and documents; and the main assessment offers a picture of the general literary (i.e., prose, document and quantitative literary) of the adults who passed the core literacy tasks.

Background Questionnaire. The 2003 NAAL Background Questionnaire collected data in a variety of background categories; it obtained valuable background information not collected in the 1992 survey. The questionnaire served three purposes:

SURVEY OF A SAMPLE OF ADULTS LIVING IN HOUSEHOLDS OR PRISONS:

Assesses literacy skills:

- Prose
- Document
- Quantitative

Collects background data on:

- Demographics
- Education
- Labor Market Experiences
- Income
- Activities

- to provide descriptive data on respondents;
- to enhance understanding of the factors that are associated with literacy skills used at home, at work, or in the community; and
- to allow for the reporting of changes over time.

The questionnaire was orally administered to every participant by an interviewer who used a computer-assisted personal interview (CAPI) system. Unlike the 1992 NALS, in which the background questions were read aloud from a printed questionnaire, in 2003, interviewers read the questions from laptop computer screens and entered the responses directly into the computer. CAPI then selected the next question based on responses to prior questions. Because the questions were targeted, a respondent did not answer all of the background questions (i.e., inapplicable questions were skipped). The questionnaire took about 28 minutes to complete.

The background questionnaire used in SAAL was the same as that used in NAAL. However, a separate questionnaire was administered for the prison component in order to address issues of particular relevance to the prison population.

Prison Component. The 2003 NAAL Prison component assesses the literacy skills and proficiencies of the U.S. adult prison population. In the 2003 assessment, approximately 1,200 adults participated, from 107 prisons (including 12 federal prisons) in 31 states.

Key features:

- provides demographic and performance data for the prison population, in comparison with the main NAAL household study of the general adult population;
- reports results that are useful to policymakers and practitioners concerned with literacy and education in correctional settings; and
- guides corrections and education professionals in the development of more effective literacy and adult education programs for prison inmates.

The principal aim of the 2003 NAAL prison component is to provide comprehensive information on the literacy and background of the U.S. adult prison population to policymakers and practitioners in order to enhance adult education in our nation's prisons and improve incarcerated adults' ability to function and

achieve their goals in the general society, in the workplace, at home, and in the community—upon their release from prison.

State Assessment of Adult Literacy (SAAL). The SAAL is an assessment of adult literacy within a participating state. Conducted in conjunction with the 2003 NAAL data collection, SAAL collected additional data within the six participating states: Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma.

Key features:

SAAL provides participating states with individually-tailored reports that offer:

- more in-depth analysis of a state's literacy, by augmenting the state's sample with the national sample;
- state and national comparisons;
- expanded background information on population groups;
- state-level scoring for FAN, ALSA, and the Health Literacy Component;
- estimates by demographic and other characteristics of interest; and
- trend data (for New York), because it participated in both the 1992 and 2003 assessments.

Health Literacy Component. The 2003 NAAL is the first large-scale national assessment in the United States to contain a component designed specifically to measure health literacy—the ability to use literacy skills to read and understand written health-related information encountered in everyday life. The Health Literacy Component establishes a baseline against which to measure progress in health literacy in future assessments.

The NAAL health literacy report—[*The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy*](#) (Kutner et al. 2007)—provides first-hand information on the status of the health literacy of American adults age 16 and older. Results are reported in terms of the four literacy performance levels—below basic, basic, intermediate, and proficient—with examples of the types of health literacy tasks that adults at each level may be able to perform.

Key features:

- reports on the health literacy skills of target audiences;
- sheds light on the relationship between health literacy and background variables, such as educational attainment, age, race/ethnicity, adults' sources of information about health issues, and health insurance coverage;
- examines how health literacy is related to prose, document, and quantitative literacy;
- provides information that may be useful in the development of effective policies and customized programs that address deficiencies in health literacy skills; and
- guides the development of health information tailored to the strengths and weaknesses of target audiences.

Fluency Addition to NAAL (FAN). FAN examines components of oral reading fluency that the main NAAL does not assess. Using speech-recognition software, FAN measures adults' ability to decode, recognize words, and read with fluency.

Key features:

- establishes a basic reading skills scale;
- identifies, for the first time, the relationship between basic reading skills and selected background characteristics, as well as performance on the main NAAL, Health Literacy Component, and prison component; and
- provides a baseline for measuring future changes in the levels and distribution of oral fluency over time.

Ultimately, FAN can improve our understanding of the skill differences between adults who are able to perform relatively challenging tasks and adults who lack basic reading skills. Such information will prove most useful to researchers, practitioners, and policymakers. For instance, adult education providers can use FAN results to develop and offer instruction and courseware that will better address the skill sets of the least literate adults. Likewise, policymakers can use FAN results to support the creation and improvement of programs serving adults with lower literacy skills.

Adult Literacy Supplemental Assessment (ALSA).

Low levels of literacy are likely to limit life chances and may be related to social welfare issues, including poverty, incarceration, and preventive health care. Given this, it has become increasingly important for researchers, policymakers, and practitioners to understand the literacy skills and deficits of the least literate adults.

ALSA is designed to assess the basic reading skills of the least literate adults. The 1992 NALS lacked a similar component. Because the least literate adults were unable to complete the 1992 assessment due to literacy-related complications (e.g., difficulty reading and writing in English; mental or learning disabilities), the 1992 NALS provided little information on these respondents.

Key features:

- enhances our understanding of the basic reading skills of the least-literate adults;
- identifies relationships between ALSA scores and selected background characteristics of adults;
- reports results for appropriate demographic groups (e.g., Black, Hispanic, and other racial/ethnic groups; ESL adults; the prison population);
- describes relationships between the performance of ALSA participants and main NAAL participants on the FAN oral reading tasks; and
- provides a baseline for measuring future changes in the levels and distribution of the least literate adults' basic reading skills over time.

Participants who scored low on the core screening questions (see "[Assessment Design](#)") were given ALSA instead of the main assessment.

The Main Assessment. NAAL main assessment reports a separate score for each of three literacy areas: prose literacy, document literacy, and quantitative literacy.

Prose literacy refers to the knowledge and skills needed to perform prose tasks—that is, to search, comprehend, and use continuous texts. Prose examples include editorials, news stories, brochures, and instructional materials.

Document literacy refers to knowledge and skills needed to perform document tasks—that is, to search, comprehend, and use continuous texts. Document examples include job applications, payroll forms, transportation schedules, maps, tables, and drug or food labels.

Quantitative literacy refers to the knowledge and skills needed to perform prose tasks—that are, to identify and perform computations, either alone or sequentially, using numbers embedded in printed materials. Examples include balancing a check book, computing a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

Periodicity

The 2003 NAAL results are comparable to those of the 1992 NALS, and for young adults 21 to 25 years old, to the 1985 young adult literacy assessment.

2. USES OF DATA

NAAL 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 1992 NALS;
- interpret the findings in light of information-processing skills and strategies, so as to inform curriculum decisions concerning adult education and training; and
- increase our understanding of the skills and knowledge associated with living in a technological society.

3. KEY CONCEPTS

NAAL is designed to measure functional English literacy. The assessment measures how adults use

printed and written information to adequately function at home, in the workplace, and in the community.

Since adults use different kinds of printed and written materials in their daily lives, NAAL measures three types of literacy—prose, document, and quantitative—and reports a separate scale score for each of these three areas. By measuring literacy along three scales, instead of just one, NAAL can provide more comprehensive data on literacy tasks and literacy skills associated with the broad range of printed and written materials adults use.

Prose Literacy

The prose literacy scale measures the knowledge and skills needed to perform prose tasks (i.e., to search, comprehend, and use continuous texts). Examples include editorials, news stories, brochures, and instructional materials.

Document Literacy

The document literacy scale measures the knowledge and skills needed to perform document tasks (i.e., to search, comprehend, and use non-continuous texts in various formats). Examples include job applications, payroll forms, transportation schedules, maps, tables, and drug or food labels.

Quantitative Literacy

The quantitative literacy scale measures the knowledge and skills required to perform quantitative tasks (i.e., to identify and perform computations, either alone or sequentially, using numbers embedded in printed materials). Examples include balancing a checkbook, figuring out a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

In addition to the prose, document, and quantitative literacy scales, the 2003 assessment included a health literacy scale. The health literacy scale contains prose, document, and quantitative items with health-related content. The items fall into three areas: clinical, prevention, and navigation of the health system.

4. SURVEY DESIGN

Data collection for the main NAAL study and the concurrent state assessment, SAAL, was conducted in 2003 using in-person household interviews. Over 18,000 adults participated, selected from a sample of over 35,000 households that represented the entire U.S. household population age 16 and over—about 222 million Americans (U.S. Census Bureau, Current Population Survey 2003). In addition, approximately 1,200 inmates from 110 federal and state prisons were

assessed in early 2004 for the prison component, which provides separate estimates of literacy levels for the incarcerated population.

All household participants received an incentive payment of \$30 in an effort to increase both the representativeness of the sample and the response rate. Black and Hispanic households were oversampled at the national level to ensure reliable estimates of their literacy proficiencies. Special accommodations were made for adults with disabilities or with limited English proficiency.

Target Population

The target population for the national household sample consisted of adults 16 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. The target population for the prison sample consisted of inmates age 16 and older in state and 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

The 2003 NAAL included two samples: (1) adults age 16 and older living in households (99 percent of the entire NAAL sample, weighted); and (2) inmates age 16 and older in state and federal prisons (1 percent of the entire NAAL sample, weighted). Each sample was weighted to represent its share of the total population of the United States, and the samples were combined for reporting.

Household sample. The 2003 NAAL household sample included a nationally representative probability sample of 35,000 households. The household sample was selected on the basis of a four-stage, stratified area sample: (1) primary sampling units (PSUs) consisting of counties or groups of contiguous counties; (2) secondary sampling units (referred to as segments) consisting of area blocks; (3) housing units containing households; and (4) eligible persons within households. Person-level data were collected through a screener, a background questionnaire, the literacy assessment, and the oral module.

Six states—Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma—purchased additional cases in their states to allow reporting at the state level. A single area sample was selected for the national NAAL sample, and additional samples were selected for the six states participating in the SAAL.

For each sample, the usual procedures for area sampling were followed: a stratified probability proportionate to size design was used for the first two stages, and systematic random samples were drawn in the last two stages.

A key feature of the national NAAL sample was the oversampling of Black and Hispanic adults, which was accomplished by oversampling segments with high concentrations of these groups. The SAAL samples did not include any oversampling of Black, Hispanic, or other racial/ethnic groups.

Although integrating the NAAL and SAAL samples at the design stage would have been more effective statistically, the states agreed to participate after the NAAL sample design and selection process had been finalized. Therefore, the approach used in the 1992 NALS was followed: selecting the SAAL samples independently of the NAAL sample and combining the samples at the estimation phase by using composite estimation.

Stage one sampling. The first stage of sampling was the selection of PSUs, which consisted of counties or groups of counties. PSUs were formed within state boundaries, which gave an improved sample for state-level estimation. One PSU was selected per stratum by using probabilities proportionate to their population within households, except in Maryland and Massachusetts, where samples of segments were selected as the first-stage units. One hundred PSUs were selected for the national sample, and 54 PSUs were selected in Kentucky, Missouri, New York, and Oklahoma. Maryland and Massachusetts had too few PSUs from which to sample; therefore, segments were selected in the first stage of sampling. After selecting the segments, 20 area clusters (quasi-PSUs) were created for Maryland and Massachusetts by grouping the selected segments into 20 geographically clustered areas to facilitate a cost-efficient approach to data collection. The true first-stage sample size is much larger because a total of 323 first-stage units (i.e., segments) were selected in Maryland and Massachusetts. Fourteen PSUs were selected for both the national NAAL and the SAAL samples; hence, the sample included a combined total of 160 unique PSUs.

Stage two sampling. In the second stage of sampling, segments (census blocks or groups of blocks) within the PSUs were selected with a probability proportionate to size; the measure of size for a segment was a function of the number of year-round housing units within the segment. In the national sample, the Black and Hispanic populations were sampled at a higher rate than the remainder of the population to

increase their sample size, whereas the state samples used no oversampling. Oversampling in the national sample was accomplished by oversampling the segments in which Black and Hispanic adults accounted for 25 percent or more of the population. There were 2,000 segments selected for the national sample and 861 segments selected across the SAAL samples, with a total of 2,800 unique segments selected across the national and six SAAL samples. (Two segments were selected for both the NAAL and SAAL samples.)

Stage three sampling. In the third stage of sampling, housing units were selected with equal probability within each segment, except for White households within high percentage of Black, Hispanic, and other race/ethnicity segments in the national component. These national sample households were subsampled after screening so that the sampling rates for White persons would be about the same in the high percentage of Black, Hispanic, and other race/ethnicity segments as in other segments. The overall sample size of housing units took into account expected losses owing to vacant housing units, units that were not housing units, and expected response rates.

Stage four sampling. The fourth stage of selection involved listing the age-eligible household members (age 16 and older) for each selected household. Subsequently, one person was selected at random within households with three or fewer eligible persons, and two persons were selected if the household had four or more eligible persons. The listing and selection of persons within households were performed with the CAPI system.

Of the 35,000 sampled households, 4,700 were either vacant or not a dwelling unit, resulting in a sample of 31,000 households. A total of 25,000 households completed the screener, which was used to select survey respondents. The final screener response rate was 81 percent (weighted).

On the basis of the screener data, 24,000 respondents age 16 and older were selected to complete the background questionnaire and the assessment; 18,000 actually completed the background questionnaire. Of the 5,500 respondents who did not complete the background questionnaire, 360 were unable to do so because of a literacy-related barrier, either the inability to communicate in English or Spanish (the two languages in which the background questionnaire was administered) or a mental disability.

The final response rate for the background questionnaire—which included respondents who completed the background questionnaire and

respondents who were unable to complete the background questionnaire because of language problems or a mental disability—was 77 percent (weighted). Of the 18,000 adults age 16 and older who completed the background questionnaire, 17,000 completed at least one question on each of the three scales—prose, document, and quantitative—measured in the adult literacy assessment. An additional 149 were unable to answer at least one question on each of the three scales for literacy-related reasons or a mental disability. The final response rate for the literacy assessment—which included respondents who answered at least one question on each scale plus the 150 respondents who were unable to do so because of language problems or a mental disability—was 97 percent (weighted).

Cases were considered complete if the respondent completed the background questionnaire and at least one question on each of the three scales or if the respondent was unable to answer any questions because of language issues (an inability to communicate in English or Spanish) or a mental disability. All other cases that did not include a complete screener, a background questionnaire, and responses to at least one question on each of the three literacy scales were considered incomplete or missing. Before imputation, the overall response rate for the household sample was 60 percent (weighted).

Imputation for nonresponse. For respondents who did not complete any literacy tasks on any scale, no information is available about their performance. Completely omitting these individuals from the analyses would have resulted in unknown biases in estimates of the literacy skills of the national population because refusals cannot be assumed to have occurred randomly. For 860 respondents who answered the background questionnaire but refused to complete the assessment for reasons other than language issues or a mental disability, regression-based imputation procedures were applied to impute responses to one assessment item on each scale by using the NAAL background data on age, gender, race/ethnicity, education level, country of birth, census region, and metropolitan statistical area status.

On the prose and quantitative scales, a response was imputed for the easiest task on each scale. On the document scale, a response was imputed for the second easiest task because that task was also included on the health literacy scale. In each of the logistic regression models, the estimated regression coefficients were used to predict missing values of the item to be imputed. For each nonrespondent, the probability of answering the item correctly was computed and then compared with a

randomly generated number between 0 and 1. If the probability of getting a correct answer was greater than the random number, the imputed value for the item was 1 (correct); otherwise, it was 0 (wrong). In addition, a wrong response on each scale was imputed for 65 respondents who started to answer the assessment, but were unable to answer at least one question on each scale because of language issues or a mental disability.

The final household reporting sample—including the imputed cases—consisted of 18,000 respondents. These 18,000 respondents include the 17,000 respondents who completed the background questionnaire and the assessment; the 860 respondents who completed the background questionnaire, but refused to do the assessment for non-literacy-related reasons (and have imputed responses to one item on each scale); and the 70 respondents who started to answer the assessment items, but were unable to answer at least one question on each scale because of language issues or a mental disability. After including the cases for which responses to the assessment questions were imputed, the weighted response rate for the household sample was 62 percent (18,000 cases with complete or imputed data and an additional 440 cases that had no assessment data because of language issues or a mental disability).

Prison sample. The 2003 assessment also included a nationally representative probability sample of inmates in state and federal prisons. The target population for the prison sample consisted of inmates age 16 and older from state and federal prisons in the United States. The sampling frame was created primarily from two data sources: the Bureau of Justice Statistics 2000 Census of State and Federal Adult Correctional Facilities (referred to in the following text as the Prison Census) and the 2003 Directory of Correctional Facilities of the American Correctional Association (ACA).

The facility universe for the NAAL Prison Component was consistent with the Prison Census. As defined for the Prison Census, the 2003 NAAL target population included the following types of state and federal adult correctional facilities: prisons; prison farms; reception, diagnostic, and classification centers; road camps; forestry and conservation camps; youthful offender facilities (except in California); vocational training facilities; drug and alcohol treatment facilities; and state-operated local detention facilities in Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont. Facilities were included in the NAAL Prison Component if they were:

- staffed with federal, state, local, or private employees;
- designed to house primarily state or federal prisoners;
- physically, functionally, and administratively separate from other facilities; and
- in operation between September 2003 and March 2004.

Specifically excluded from the NAAL Prison Component were:

- privately operated facilities that were not exclusively for state or federal inmates;
- military facilities;
- Immigration and Naturalization Service facilities;
- Bureau of Indian Affairs facilities;
- facilities operated and administered by local governments, including those housing state prisoners;
- facilities operated by the U.S. Marshals Service, including the Office of the Detention Trustee;
- hospital wings and wards reserved for state prisoners; and
- facilities housing only juvenile offenders.

Even though they contain inmates up to age 21, juvenile facilities were excluded from NAAL for two reasons: (1) to remain consistent with the facilities listed in the Prison Census; and (2) to promote cost efficiency because it would not have been cost-effective to visit these facilities to sample the small number of inmates 16 years of age and older.

Inmate sampling frames were created by interviewers at the time they visited the prisons. The frame consisted of all inmates occupying a bed the night before inmate sampling was conducted.

Approximately 110 prisons were selected to participate in the adult literacy assessment. The final prison response rate was 97 percent (weighted). Among the inmates in these prisons, 1,300 inmates ages 16 and older were randomly selected to complete the background questionnaire and assessment. Of these 1,300 selected inmates, 1,200 completed the

background questionnaire. Of the 140 inmates who did not complete the background questionnaire, about 10 were unable to do so because of a literacy-related barrier (either the inability to communicate in English or Spanish) or a mental disability.

The final response rate for the prison background questionnaire—which included respondents who completed the background questionnaire and respondents who were unable to complete the background questionnaire because of language problems or a mental disability—was 91 percent (weighted). Of the 1,200 inmates who completed the background questionnaire, 1,100 completed at least one question on each of the three scales—prose, document, and quantitative—measured in the adult literacy assessment. An additional 10 inmates were unable to answer at least one question on each of the three scales for literacy-related reasons. The final response rate for the literacy assessment—which included respondents who answered at least one question on each scale or were unable to do so because of language problems or a mental disability—was 99 percent (weighted).

The same definition of a complete case used for the household sample was also used for the prison sample, and the same rules were followed for imputation. Before imputation, the final response rate for the prison sample was 87 percent (weighted).

Imputation for nonresponse. One response on each scale was imputed on the basis of background characteristics for 30 inmates who completed the background questionnaire, but had incomplete or missing assessments for reasons that were not literacy related. The statistical imputation procedures were the same as for the household sample. The background characteristics used for the missing data imputation for the prison sample were prison security level, region of country/type of prison, age, gender, educational attainment, country of birth, race/ethnicity, and marital status. A wrong response on each scale was imputed for the inmates who started to answer the assessment, but were unable to answer at least one question on each scale because of language issues or a mental disability. The final prison reporting sample—including the imputed cases—consisted of 1,200 respondents. After the cases for which responses to the assessment questions were imputed were included, the weighted response rate for the prison sample was 88 percent (1,200 cases with complete or imputed data and an additional 20 cases that had no assessment data because of language issues or a mental disability).

Assessment Design

The NAAL interview was conducted in the order described below.

First, every respondent completed a background questionnaire that collected data on demographic, socioeconomic, and other factors associated with literacy.

Next, every respondent completed seven core screening questions, which were among the easiest in the assessment.

Similar in structure to the main NAAL assessment questions, the core questions determined whether a respondent's skills were sufficient to participate in the main NAAL assessment or if the individual should be routed to ALSA. Interviewers used a scoring rubric to code respondents' answers to each core question (e.g., "1" for correct, "2" for wrong, and "3" for no response). Interviewers entered the codes into a CAPI System, which selected respondents for ALSA using an empirically derived algorithm that predicts very low performance on the main NAAL. ALSA assessed the ability of the least literate adults to identify letters and numbers and to comprehend simple prose materials. Those participants who scored low on the basic core screening questions took ALSA instead of the main NAAL.

After completing either the main NAAL assessment booklet or ALSA, every respondent took FAN. FAN used speech-recognition software to assess adults' ability to decode and recognize words and to read with fluency.

Data Collection and Processing

Reference dates. Household data collection was conducted from March 2003 through February 2004; prison data collection was conducted from March through July 2004.

Data collection. Household interviews took place in respondents' homes; prison interviews generally took place in a classroom or library in the prison. Whenever possible, interviewers administered the background questionnaire and assessment in a private setting. Unless there were security concerns, a guard was not present in the room when inmates were interviewed.

Interviewers used a CAPI system programmed into laptop computers. The interviewers read the background questions from the computer screen and entered all responses directly into the computer. Skip patterns and follow-up probes for contradictory or out-of-range responses were programmed into the computer.

After completing the background questionnaire, respondents were handed a booklet with the assessment questions. The interviewers followed a script that introduced the assessment booklet and guided the respondent through the assessment.

Each assessment booklet began with the same seven screening questions. After the respondent completed the screening questions, the interviewer asked the respondent for the book and used an algorithm to determine, on the basis of the responses to the questions, whether the respondent should continue in the main assessment or be placed in ALSA. Three percent (weighted) and 5 percent (unweighted) of adults were placed in the ALSA.

ALSA is a performance-based assessment that allowed adults with marginal literacy to demonstrate what they could and could not do when asked to make sense of various forms of print. The ALSA started with simple identification tasks and sight words and moved to connected text, using authentic, highly contextualized material commonly found at home or in the community.

Respondents were routed to an alternative assessment (ALSA) based on their performance on the seven easy screening tasks at the beginning of the literacy assessment. Because the ALSA respondents answered most, or all, of these questions incorrectly, if they were placed on the NAAL scale, they would have been classified on the NAAL scale as below basic level on the health scale.

A respondent who continued in the main assessment was given back the assessment booklet, and the interviewer asked the respondent to complete the tasks in the booklet and guided the respondent through them. The main assessment consisted of 12 blocks of tasks with approximately 11 questions in each block, but each assessment booklet included only 3 blocks of questions. The blocks were spiraled so that across the 26 different configurations of the assessment booklet, each block was paired with every other block and each block appeared in each of the three positions (first, middle, last) in a booklet.

For ALSA interviews, the interviewer read the ALSA script from a printed booklet and classified the respondent's answers into the response categories in the printed booklet. ALSA respondents were handed the materials they were asked to read.

Following the main assessment or ALSA, all respondents were administered FAN (the oral fluency assessment). Respondents were handed a booklet with passages, number lists, letter lists, word lists, and

pseudoword lists to read orally. Respondents read into a microphone that recorded their responses on the laptop computer.

Accommodations. With the passage of the Americans with Disabilities Act and the growth of America's immigrant population, assessment programs like NAAL must consider issues of inclusion and accommodation. The 2003 NAAL provided for two types of accommodations—administrative and language.

Administrative accommodations were made for adults with disabilities. First, NAAL is inherently accommodating because the assessment was conducted one-on-one in the respondent's home. Second, all respondents with disabilities received additional time to complete the assessment, if necessary.

Language accommodations were made for adults with limited English proficiency or whose primary language is not English. Questions on the background questionnaire were available in either English or Spanish. In addition, instructions for FAN, ALSA, and the core screening test questions were given in either English or Spanish. However, the stimulus materials for these questions were in English since NAAL's main objective is to assess literacy in English.

Results are reported separately for non-native speakers of English and compared to the results of native speakers of English. Thus, the unique needs of English as a Second Language (ESL) adults may be better understood by researchers, policymakers, and practitioners.

Data processing. The NAAL assessment questions were open-ended and thus required scoring by trained scorers. NAAL experts have developed scoring rubrics that detail the rules necessary for scoring each assessment question.

In order to make NAAL scores meaningful, the scores were grouped into performance levels to provide information that could more easily be understood and used by the public and policymakers. The performance levels were developed to characterize the status of English language literacy of American adults and include the following: nonliterate in English, below basic, basic, intermediate, and proficient literacy. For reporting purposes adults classified as nonliterate in English are included in the below basic literacy level. The 2003 NAAL performance levels are different from the five levels NCES used to report NALS results in 1992. However, in order to make comparisons across years, the 1992 data were reanalyzed and the new performance levels were applied to the 1992 data.

NAAL scoring is designed to measure adults' abilities to perform literacy tasks in everyday life. Since adults are likely to make mistakes as they interact with printed and written material, NAAL scorers make allowances for partial responses and writing errors.

While most responses are either correct or incorrect, a response can be partially correct if the information provided is still useful in accomplishing the task. For example, a respondent who writes the wrong product price on a catalog order form could receive partial credit, because in real life such a minor error would not necessarily result in the placement of an incorrect order (since other information is provided, such as product name and price). However, if a respondent miswrites a social security number on a government application form, such an error would not receive partial scoring.

Similarly, responses containing writing errors—grammatical and spelling errors, use of synonyms, incomplete sentences, or circling instead of writing the correct answer—are scored as correct as long as the overall meaning is correct and the information provided accomplishes the task. However, if a respondent is filling out a form and writes the answer on the wrong line, or if, for a quantitative task, the calculation is right but the respondent writes the wrong answer in the blank, then the response is scored as incorrect.

During the task development stage, scoring experts developed scoring rubrics that detailed the rules for scoring each assessment question. To ensure that all assessment questions were scored accurately, NAAL scoring rubrics underwent several stages of verification both before and after the assessment was administered.

Before the main NAAL study began, a field test of about 1,400 adults was conducted to help identify and screen out problems with the scoring rubrics, such as alternative correct responses and scoring rubrics that are difficult to implement consistently (thus leading to low rates of interrater reliability).

After the main study ended, a sample of responses from the household and prison interviews was scored using the scoring rubrics. As the test developers scored the sample responses, they made adjustments to the scoring rubrics to reflect the kinds of responses adults gave during the assessment. Together, these sample responses and the revised scoring rubrics were used in training the scorers who scored the entire assessment.

In a group setting, scorers were trained to recognize each task and its corresponding scoring rubric, as well as sample responses that are representative of correct, partially correct, and incorrect answers. After group

training, readers scored numerous practice questions before they began to score actual booklets.

To ensure that readers were scoring accurately, 50 percent of the assessment questions were subject to a second interrater reliability check, in which a second reader scored the booklet and the scores of the first and second readers were compared. Interrater reliability is the percentage of times two readers agree exactly in their scores. (In 1992, the average percentage of agreement was 97 percent.) Any batch of questions that exceeded a low level of scoring mistakes was sent back to the scorers for corrections. Also, the scoring supervisor discussed the discrepancy with the scorers involved. Quality control procedures like this ensured reliability of the scoring.

Performance levels. Performance levels are important because they provide the ability to group people with similar literacy scores into a relatively small number of categories of importance to the adult education community, much like grouping students with similar scores on a test into various letter grades (e.g., A or B). A benefit of having performance levels is that they enable NAAL to characterize American adults' relative literacy strengths and weaknesses by describing the nature and difficulty of the literacy tasks that participants at each level can perform with a reasonably high rate of success.

Performance levels were determined in response to a request from NCES to the National Research Council (NRC), which convened a Committee on Performance Levels for Adult Literacy. The committee's goal was to do the following in an open and public way: evaluate the literacy levels used by NAAL's 1992 predecessor survey, and recommend a set of performance levels that could be used in reporting the 2003 results and also be applied to the 1992 results in order to make comparisons across years.

New performance levels. After reviewing information about the 1992 and 2003 assessments as well as feedback from stakeholders (e.g., adult literacy practitioners), the NRC committee specified a new set of performance levels intended to correspond to four policy-relevant categories of adults, including adults in need of basic adult literacy services. The next step was to determine the score ranges to be included in each level for each of the three NAAL literacy scales—prose, document, and quantitative literacy.

Score ranges. To determine the score ranges for each level, the committee decided to use the "bookmark" method. Initial implementation of the method involved describing the literacy skills of adults in the four policy-relevant levels, and holding two sessions with

separate panels of “judges” consisting of adult literacy practitioners, officials with state offices of adult education, and others. One group of judges focused on the 1992 assessment tasks and the other group focused on the 2003 assessment tasks.

Bookmarks. For each literacy area (prose, document, and quantitative), the judges were given, in addition to descriptions of the performance levels, a booklet of assessment tasks arranged from easiest to hardest. The judges’ job was to place “bookmarks” in the set of tasks that adults at each level were “likely” to get right. The term “likely” was defined as “67 percent of the time,” or two out of three times, and statistical procedures were used to determine the score associated with a 67 percent probability of performing the task correctly. The bookmarks designated by the judges at the two sessions were combined to produce a single bookmark-based cut score for each performance level on each of the three literacy scales.

Quasi-contrasting groups approach. To refine the bookmark-based cut scores, which indicated the lowest score to be included in each performance level, the committee used a procedure it termed the “quasi-contrasting groups approach.” The committee compared the 2003 bookmark-based cut scores with the 1992 scores associated with various background variables, such as educational attainment. The criterion for selecting the background variables was potential usefulness for distinguishing between adjacent performance levels, such as basic and below basic (e.g., having some high school education vs. none at all; reporting that one reads well vs. not well; reading a newspaper sometimes vs. never reading a newspaper; reading at work sometimes or more often vs. never reading at work).

In each case, the midpoint between the average scores of the two adjacent performance levels (below basic and basic; basic and intermediate; intermediate and proficient) was calculated and averaged across the variables that provided contrast between the groups. The committee developed a set of rules and procedures for deciding when and how to make adjustments to the bookmark cut scores when the cut scores associated with the selected background variables were different from the bookmark-based scores.

Nonliterate in English classification. The NRC committee recommended that NCES distinguish a fifth group of adults with special importance to literacy policy—those who are nonliterate in English. As originally defined by the committee, this category consisted of adults who performed poorly on a set of easy screening tasks in 2003 and therefore were routed

to an alternative assessment for the least literate adults (i.e., ALSA). Because the 1992 assessment included neither the alternative assessment nor the 2003 screening tasks, adults in this category cannot be identified for 1992.

To provide a more complete representation of the adult population that is nonliterate in English, NCES expanded the category to include not only the 3 percent of adults who took the alternative assessment, but also the 2 percent who were unable to be tested at all because they knew neither English nor Spanish (the other language spoken by interviewers). Thus, as defined by NCES, the category included about 5 percent of adults in 2003.

Refinements made before using the new levels. The new performance levels were presented to NCES as recommendations. Having accepted the general recommendations, NCES incorporated a few refinements before using the levels to report results. First, NCES changed the label of the top category from advanced to proficient because the term “proficient” better conveys how well the upper category of adults performs. Second, NCES added sample tasks from the 2003 assessment to illustrate the full range of tasks that adults at each level can perform, as well as a brief (one-sentence) summary description for each level to enhance public understanding. Third, as outlined in the previous paragraph, NCES included additional adults in the “nonliterate in English” category.

Estimation Methods

Weighting. As discussed above, NAAL included both a household sample and a prison sample. The household sample was further divided into the cases selected for the national sample and the additional cases selected in the six SAAL states. Weighting was done separately for the household and prison samples. However, the weights were developed so that the two samples could be used together in a combined sample.

Household sample weighting. Differential probabilities of selection into the NAAL household sample were adjusted by computing base weights for all adults selected into the sample. The base weight was calculated as the reciprocal of a respondent’s final probability of selection. The weights were adjusted for nonresponse at both the screener level and the background questionnaire level. Additionally, trimming procedures were followed to reduce the impact of extreme weights. The background questionnaire weighting steps were done separately for the national and SAAL household samples, and each sample was calibrated separately to population estimates based on 2003 Current Population Survey

(CPS) data. To combine the NAAL and SAAL household samples, composite weights were calculated for the respondents in the six participating SAAL states and the respondents in the national NAAL household sample in these six states. The composite weights were adjusted through poststratification and raking to match the 2003 CPS data.

Prison sample weighting. The prison component weighting consisted of four main steps. First, prison base weights were constructed using the probability of selection for each prison into the sample. Then, a nonresponse adjustment was made to the prison base weights to account for nonparticipating prisons. Next, inmate base weights were calculated using the prison nonresponse-adjusted weight and the within-prison sampling rate. Finally, the inmate base weights were raked to Bureau of Justice Statistics control totals to account for inmate nonresponse and noncoverage.

Variance estimation. A complex sample design was used to select assessment respondents. The properties of a sample selected through a complex design can be very different from those of a simple random sample. (In a simple random sample, every individual in the target population has an equal chance of selection and the observations from different sampled individuals can be considered to be statistically independent of one another.) Sampling weights should be used to account for the fact that the probabilities of selection were not identical for all respondents. All population and subpopulation characteristics based on the NAAL data should use sampling weights in their estimation.

Since the respondents were selected using complex sample design, conventional formulas for estimating sampling variability that assume simple random sampling (and, hence, independence of observations) are inappropriate. Standard errors calculated as though the data had been collected from a simple random sample would generally underestimate sampling errors. Therefore, the properties of the complex data collection design should be taken into account during the analysis of the data.

Scaling. Each respondent to NAAL received a booklet that included 3 of the 13 assessments blocks. Because each respondent did not answer all of the NAAL items, item response theory (IRT) methods were used to estimate average scores on the health, prose, document, and quantitative literacy scales; a simple average percent correct would not allow reporting results that were comparable for all respondents. IRT models calculate the probability of answering a question correctly as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a

common scale on which performance on some latent trait can be compared across groups, such as those defined by sex, race/ethnicity, or place of birth.

IRT models assume that an examinee's performance on each item reflects characteristics of the item and characteristics of the examinee. All models assume that all items on a scale measure a common latent ability or proficiency dimension (e.g., prose literacy) and that the probability of a correct response on an item is uncorrelated with the probability of a correct response on another item, given fixed values of the latent trait. Items are measured in terms of their difficulty as well as their ability to discriminate among examinees of varying ability.

The assessment used two types of IRT models to estimate scale scores. The two-parameter logistic (2PL) model was used for dichotomous items (that is, items that are scored either right or wrong). For the partial credit items, the graded response logistic (GRL) model was used. The scale indeterminacy was solved by setting an origin and unit size to the reported scale means and standard deviations from the 1992 assessment. Linear transformation was performed to transform the original scale metric to the final reporting metric.

IRT models predict the probability of success on an item for each point along the latent ability scale. By selecting a criterion value for this probability, a single scale point can be associated with the difficulty of each item, and visual displays can be constructed showing the difficulty of selected items along the scale. Such item maps aid in interpreting the assessment scales and in describing the performance levels. The assessment conformed to common industry practice by choosing the value of 0.67 as its response probability convention.

5. DATA QUALITY AND COMPARABILITY

The NAAL sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest.

Sampling Error

In the 2003 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 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 2003 NAAL 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 age 16 and older on the lists from which the sampled respondents were drawn. Special procedures and edits were built into NAAL 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 2003 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—the 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 the 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. Thus, response rates were calculated for each of the three instruments for the household samples. 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.

For occupied households, “refusal or breakoff” was the most common explanation for nonresponse to the screener and the 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 not 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.

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 levels 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.

The household sample was subject to unit nonresponse from the screener, background questionnaire, literacy assessment, and oral module and to item nonresponse to background questionnaire items. Although all

Table NAAL-1. Weighted and unweighted unit response rates in the household sample of the National Assessment of Adult Literacy, by survey component: 2003

Component	Weighted response rate (percent)	Unweighted response rate (percent)
Screener	81.2	81.8
Background questionnaire	76.6	78.1
Literacy assessment	96.6	97.2
Overall response rate before imputation	60.1	62.1
Overall response rate after imputation	62.1	63.9

SOURCE: Greenberg, E., and Jin, Y. (2007). *2003 National Assessment of Adult Literacy: Public-Use Data File User's Guide* (NCES 2007-464). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

background questionnaire items had response rates of more than 85 percent, two stages of data collection—the screener and the background questionnaire—had unit response rates below 85 percent and thus required an analysis of the potential for nonresponse bias.

Table NAAL-1 presents a summary of the household response rate and table 14 presents a summary of the prison response rate.

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 CD-ROM: 2003 National Assessment of Adult Literacy Public-Use Data File User's Guide (Greenberg & Jin 2007) 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.

Table NAAL-2. Weighted and unweighted response rates in the prison sample of the National Assessment of Adult Literacy, by survey component: 2003

Component	Weighted response rate (percent)	Unweighted response rate (percent)
Prison	97.3	97.3
Background questionnaire	90.6	90.4
Literacy assessment	98.9	98.8
Overall response rate before imputation	87.2	86.8
Overall response rate after imputation	88.3	87.9

SOURCE: Greenberg, E., and Jin, Y. (2007). *2003 National Assessment of Adult Literacy: Public-Use Data File User's Guide* (NCES 2007-464). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Nonresponse bias. NCES statistical standards require a nonresponse bias analysis when the unit response rate for a sample is less than 85 percent. The nonresponse bias analysis of the household sample revealed differences in the background characteristics of respondents who participated in the assessment compared with those who refused.

In bivariate unit-level analyses at the screener and background questionnaire stages, estimated percentages for respondents were compared with those for the total eligible sample to identify any potential bias owing to nonresponse. Although some statistically significant differences existed, the potential for bias was small because the absolute difference between estimated percentages was less than 2 percent for all domains considered. Multivariate analyses were conducted to further explore the potential for nonresponse bias by identifying the domains with the most differential response rates. These analyses revealed that the lowest response rates for the screener were among dwelling units in segments with high median income, small average household size, and a large proportion of renters. The lowest response rates for the background questionnaire were among males age 30 and older in segments with high median income.

However, the variables used to define these areas and other pockets with low response rates were used in weighting adjustments. The analysis showed that weighting adjustments were highly effective in reducing the bias. The general conclusion was that the potential amount of nonresponse bias attributable to unit nonresponse at the screener and background questionnaire stages was likely to be negligible.

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

6. CONTACT INFORMATION

For content information about the NAAL project, contact:

Andrew Kolstad
Phone: (202) 502-7374
E-mail: 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

Greenberg, E., and Jin, Y. (2007). *2003 National Assessment of Adult Literacy: Public-Use Data File User's Guide* (NCES 2007-464). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., and Dunleavy, E. (2007). *Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy* (NCES 2007-480). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kutner, M., Greenberg, E., Jin, Y., and Paulsen, C. (2006). *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy* (NCES 2006-483). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

White, S., and Dillow, S. (2006). *Key Concepts and Features of the 2003 National Assessment of Adult Literacy* (2006-471). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.