

## **Chapter 1**

### **THE NATIONAL ADULT LITERACY SURVEY: AN OVERVIEW**

Lynn Jenkins, Wordsworth Writing and Editing (formerly with Educational Testing Service)  
Stéphane Baldi, American Institutes for Research

#### **1.1 INTRODUCTION**

The Adult Education Amendments of 1988 required the U.S. Department of Education to submit a report to Congress defining literacy and measuring the nature and extent of literacy among adults in the nation. To satisfy these requirements, the National Center for Education Statistics (NCES) and the Division of Adult Education and Literacy planned a nationally representative household survey to assess the literacy skills of the adult population in the United States. In September 1989, NCES awarded a four-year contract for that purpose to Educational Testing Service (ETS) with a subcontract to Westat, Inc., for sampling and field operations.

The National Adult Literacy Survey is the third and largest assessment of adult literacy funded by the Federal government and conducted by ETS. The two previous efforts included a 1985 household survey of the literacy skills of 21- to 25-year-olds, funded by the U.S. Department of Education, and a 1989-90 survey of the literacy proficiencies of job seekers, funded by the U.S. Department of Labor.

In 1992, nearly 13,600 individuals age 16 and older, randomly selected to represent the adult population in this country, were surveyed in their homes. In addition, about 1,000 randomly selected adults age 16 through 65 were surveyed in each of 11 states that chose to participate in a concurrent State Adult Literacy Survey designed to produce state-level results comparable to the national data. In addition to the household samples, 1,147 inmates from 87 state and Federal prisons were randomly surveyed to represent the inmate population in the United States. Their participation helped to provide better estimates of the literacy levels of the total population and made it possible to report on the literacy proficiencies of this important segment of society.

Each individual who participated in the National and State Adult Literacy Surveys was asked to provide background demographic information and to complete a booklet of literacy tasks. These tasks were carefully constructed to measure respondents' ability to read and use a wide array of printed and written materials.

The survey results comprise an enormous set of data that includes more than a million responses to the literacy tasks and background questions. More important than the size of the database, however, is the fact that it provides information that is essential to understanding this nation's literacy resources. Specifically, the National Adult Literacy Survey data give policy makers, business and labor leaders,

educators, researchers, and citizens vital information on the condition of literacy in the United States. 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, including those targeted as “at risk;”
- Characterize adults’ literacy skills in terms of demographic and background information (such as reading characteristics, education, and employment experiences);
- Profile the literacy skills of the nation’s work force;
- Compare assessment results from the current study with those from the 1985 literacy survey of young adults;
- 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.

This chapter describes the design for the 1992 National Adult Literacy Survey and gives an overview of the steps involved in its implementation, from the development of a working definition of literacy to the creation of edited data files. The major components of the implementation of the survey are presented here as a tool to help the reader gain an overview of the National Adult Literacy Survey without having to read each individual chapter. For more detailed or technical information, the reader is referred to the specific chapters of this technical report as well as to the booklet *Assessing Literacy* (Campbell, Kirsch, and Kolstad, 1992) and the initial report on the survey, *Adult Literacy in America* (Kirsch, Jungeblut, Jenkins, and Kolstad, 1993).

The organization of this chapter is as follows:

Section 1.2 provides an overview of the development of the working definition of literacy that underlies the National Adult Literacy Survey.

Section 1.3 summarizes the stratified random sampling procedures used for the national, state, and prison components of the survey.

Section 1.4 gives an overview of the use and computation of weights used in the 1992 National Adult Literacy Survey to permit inferences from persons included in the sample to the populations from which they were drawn.

Section 1.5 discusses the development of cognitive and background questions in the survey instrument.

Section 1.6 summarizes the field operations and data collection in the household and prison surveys.

Section 1.7 describes the data processing operations, including data entry, validation, the treatment of missing data, and the creation of edited data files.

Section 1.8 discusses the Item Response Theory (IRT) scaling model and the plausible values methodology used to score respondents’ performance to the items in the questionnaire.

Section 1.9 discusses the establishment of literacy levels for the National Adult Literacy Survey.

## **1.2 DEFINING LITERACY**

Although few would deny the importance of literacy in today's society, a shared belief in the value of literacy does not imply consensus on how to define and measure it. In fact, there are widely varying opinions about the skills that individuals need to function successfully in their work, in their personal lives, and in society, and about the ways in which these skills should be assessed. As a result, there have been widely conflicting diagnoses of the literacy problem in this country.

A committee of experts from business and industry, labor, government, research, and adult education worked with ETS staff to develop the definition of literacy that underlies the National Adult Literacy Survey, 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.

Drawing on the two earlier studies of adult literacy conducted by ETS and funded by the Federal government (Kirsch and Jungeblut, 1986; Kirsch, Jungeblut, and Campbell, 1992), the Literacy Definition Committee rejected the types of arbitrary standards—such as signing one's name, completing five years of school, or scoring at a particular grade level on a school-based measure of reading achievement—that have long been used to make judgments about adults' literacy skills. Through a consensus process, the committee adopted the following definition of literacy, initially developed for the 1985 young adult survey:

*Using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.*

This definition of literacy extends beyond simple decoding and comprehension to include a broad range of skills that adults use in accomplishing many different types of literacy tasks associated with work, home, and community contexts.

## **1.3 THE SAMPLE**

The National Adult Literacy Survey was administered to three samples: 1) a national household sample, 2) household samples from 11 states, and 3) a national sample of prison inmates. Both the national and state household samples were based on four-stage, stratified sampling. The prison sample was based on two-stage sampling. While the national and state household samples were drawn using the same sampling strategy, they differed in two ways: blacks and Hispanics were oversampled only in the national sample, and the target population for the national sample consisted of adults age 16 or older while for the state sample the target population consisted of adults ages 16 to 64. Blacks and Hispanics were oversampled in the national sample based on the key objective of the national sample: to provide reliable statistics for the adult population along with the prespecified domains. The prespecified domains included a racial/ethnic domain

and an adults aged 65 and older domain. While the states wanted reliable statics, they were not concerned with the specific domains, and thus did not oversample them.

The four sampling stages for the national and state samples were: (1) the selection of primary sampling units (PSUs) consisting of counties or groups of counties, (2) the selection of segments consisting of census blocks or groups of blocks, (3) the selection of households, and (4) the selection of age-eligible individuals. In the first stage of sampling, the PSUs were stratified according to census region, metropolitan status, percentage of black residents, percentage of Hispanic residents, and, whenever possible, per capita income. In the second stage of sampling, census blocks or groups of blocks within each PSU were selected with a probability proportional to the number of housing units. In the third stage, a list of all housing units was created. A list of all housing units within the boundaries of each segment was then selected. Households were selected with equal probability within each segment of census blocks or groups of blocks, except for White, non-Hispanic households in high-minority segments in the national component. Finally, in the fourth stage of sampling, one person was randomly selected from each household with fewer than four eligible members and two persons were randomly selected from each household with four or more eligible members, from a list of all age-eligible household members (age 16 or older for the national sample and age 16 to 64 for the state samples). The same stratification methods, PSU construction, sample design and instruments were used for both the national and state designs.

In addition, at the request of the Office of Management and Budget, a subsample of 1,812 households drawn from the 2,064 segments in the national sample was randomly selected following the steps outlined above in order to yield approximately 1,000 respondents who would be administered the survey without a \$20 incentive. This was done to be able to compare the incentive versus non-incentive response rates as well as assess the effect of incentives on response patterns.

For the prison survey, the two sampling stages were (1) the selection of primary sampling units (PSUs), and (2) the selection of inmates within each PSU. In this case, PSUs consisted of state or Federal adult correctional facilities, which were selected with a probability proportional to size. In the second stage, inmates were selected with a probability inversely proportional to the number of inmates, up to 22 inmates in a facility. Chapter 2 provides a discussion of the sample design.

#### **1.4 WEIGHTING**

Whenever various subsets of the population are sampled at different rates or have different rates of selection or response, weights are necessary in order to permit inferences from persons included in the sample to the populations from which they were drawn, as well as to have sample estimates reflect estimates of the larger population. For example, in the national component of the National Adult Literacy Survey, blacks and Hispanics were oversampled to ensure reliable estimates of literacy proficiencies and to permit analyses of the performance of different subpopulations. Furthermore, because only one person was selected in

households with fewer than four eligible members, members of households with only one eligible member had twice the chance of selection as members of households with two eligible members, and three times the chance of selection as those in households with three eligible members. In such cases, weights are necessary to prevent serious bias in the estimates. Specifically, in the National Adult Literacy Survey, weights were computed to accomplish the following five objectives: (1) to permit unbiased estimates, taking account of the fact that all persons in the population did not have the same probability of selection, (2) to combine the state and national samples in an efficient manner, (3) to bring data up to the dimensions of the population totals, (4) to use auxiliary data on known population characteristics in such a way as to reduce sampling errors, and (5) to minimize biases arising from differences between cooperating and non-cooperating persons in the sample.

Differential probability of selection was corrected by computing base weights for all persons selected into the sample. For all three components (national, state, and prison), the base weight was calculated as the reciprocal of a respondent's final probability of selection. Furthermore, to combine the state and national samples, composite weights were calculated for the respondents in the 11 state samples and the respondents in the national sample PSUs in the 11 states. Finally, to adjust for non-response, weights were adjusted through post stratification and raking to match 1990 census totals. Chapter 3 provides detailed information on the weighting procedures.

### **1.5 THE SURVEY INSTRUMENT: MEASURING LITERACY**

The Literacy Definition Committee endorsed the notion that literacy is neither a single skill suited to all types of texts, nor an infinite number of skills, each associated with a given type of text or material. Rather, as suggested by the results of the young adult and job seeker surveys, an ordered set of literacy skills appears to be called into play to accomplish diverse types of tasks. Accordingly, in addition to adopting the definition of literacy that guided the earlier young adult and job-seeker studies, the Literacy Definition Committee adopted three literacy scales—prose, document, and quantitative—to report the results of the surveys.

*Prose literacy* involves the knowledge and skills needed to understand and use information from texts that include editorials, news stories, poems, and fiction; for example, finding a piece of information in a newspaper article, interpreting instructions from a warranty, inferring a theme from a poem, or contrasting views expressed in editorials.

*Document literacy* concerns the knowledge and skills required to locate and use information contained in materials that include job applications, payroll forms, transportation schedules, maps, tables, and graphs; for example, locating a particular intersection on a street map, using a schedule to choose the appropriate bus, or entering information on an application form.

*Quantitative literacy* involves the knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed materials; for example, balancing a

checkbook, figuring out a tip, completing an order form, or determining the amount of interest from a loan advertisement.

The prose, document, and quantitative scales were augmented in the current survey through the addition of new assessment tasks that took into account the following:

- Continued use of open-ended simulation tasks;
- Continued emphasis on tasks that measure a broad range of information-processing skills and cover a wide variety of contexts;
- Increased emphasis on simulation tasks that require brief written and/or oral responses;
- Increased emphasis on tasks that ask respondents to describe how they would set up and solve a problem; and
- Use of a simple, four-function calculator to solve selected quantitative problems.

Approximately 110 new assessment tasks were field tested, and 81 of these were selected for inclusion in the survey. These 81 new assessment tasks were added to a pool of 85 tasks that were administered in both the young adult and job-seeker assessments (Kirsch and Jungeblut, 1986a and 1992). Thus, the National Adult Literacy Survey consisted of a total of 166 assessment tasks. By administering a common set of assessment tasks in each of the three literacy surveys, it is possible to compare results across time and across population groups.

No individual could be expected to respond to the entire set of 166 simulation tasks administered as part of the National Adult Literacy Survey. It was therefore necessary to adopt a survey design that would give each person participating in the study a subset of the total pool of literacy tasks, while at the same time ensuring that each of the 166 tasks was administered to a nationally representative sample of the adult population. Literacy tasks were assigned to blocks or sections that could be completed in about 15 minutes, and these blocks were then compiled into booklets in such a way 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 booklets, each of which could be completed in about 45 minutes. During a personal interview, each survey participant was asked to complete one booklet.

In addition to the time allocated for the literacy tasks, approximately 20 minutes were devoted to obtaining personal information from respondents. Major areas explored included background demographics, education, labor market experiences, income, and literacy-related activities. These background data help to improve understanding of the ways in which various characteristics are associated with demonstrated literacy skills.

Trained interviewers surveyed some 13,600 adults age 16 and older, chosen to represent the household population nationwide. In addition to the national samples, approximately 1,000 adults ages 16 to 64 were assessed in each of the states that chose to participate in the State Adult Literacy Survey, a special study designed to provide state-level data comparable to the national results. California, Illinois, Indiana,

Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, and Washington conducted their surveys at the same time as the national survey. (One additional state, Florida, was surveyed at a later date.) To permit comparisons of the state and national results, the survey instruments administered to the state and national samples were identical.

Finally, 1,147 inmates from 87 state and Federal prisons were surveyed. Because some questions included in the household survey were inappropriate for the prison population, a revised version of the background questionnaire was developed that included queries about current offenses, criminal history, and prison work assignments, as well as education and work force experiences. To ensure comparability with the national survey, the simulation tasks (tasks that simulate the demands that adults encounter when they interact with printed materials on a daily basis) given to the prison participants were the same as those given to the household survey population.

A total of 26,091 adults gave, on average, over an hour of their time to complete the National Adult Literacy Survey instruments. Those who agreed to participate in the survey and completed as much of the assessment as their skills allowed were paid \$20 for their time. Responses from the national, state, and prison samples were combined to yield the best possible performance estimates. Chapter 4 describes the development of the survey instrument.

## **1.6 FIELD OPERATIONS**

Field operations and data collection for the National Adult Literacy Survey were the responsibility of Westat, Inc. The literacy survey was conducted between February and August 1992 by more than 400 trained interviewers, some of whom were bilingual in English and Spanish. All components of the survey sample were worked simultaneously, including the national sample, the state sample, and the prison sample. The field organization was headed by the survey field director, who reported directly to the Westat project director and who was supported by four home-office field managers and 24 field supervisors located across the United States. Each supervisor was supported in the field by an editor who was responsible for completely editing each case received from the field.

Interviewers were recruited directly based on Westat's computerized field personnel file containing information on over 4,000 field staff who had worked for Westat in the previous three years. A total of 456 interviewers were recruited, of which 2 did not attend training and 2 were released at training. Training consisted of a 3-day in-person training program, preceded by home study.

The administration of the national and state household surveys to respondents occurred in three overlapping stages: an initial phase, in which each area segment was assigned to an interviewer; a reassignment phase, in which incomplete interviews were given to another interviewer in the same PSU; and a special non-response conversion phase, in which the home office assembled a special traveling team of the most experienced interviewers to perform a non-response conversion effort.

For the survey of the prison population, 51 interviewers were recruited from among the household survey workforce. These interviewers received an additional 1-day, in-person training session emphasizing collecting data on criminal history and prison employment. Interviewers were required to perform a careful edit before leaving the facility because it was not possible to recontact the prisoners if errors were made.

An automated management system tracked and recorded the progress of fieldwork throughout the interview phase. In addition, progress was monitored weekly through telephone conferences between field supervisors, Westat home office staff, and ETS staff. Quality control checks were performed throughout the field data collection period and took the form of careful editing of completed documents, validation of 10 percent of each interviewer's closed-out cases, observations of interviews in person and by tape recordings, and observation of supervisors by the Westat home office and ETS staff.

As a result of the careful design of the field operations, the response rates achieved were quite favorable. Eighty-one percent of eligible respondents for the combined state and national surveys answered the background questionnaire. Of those, 95.8 percent completed the booklet of literacy exercises. For the prison population, 85.6 percent completed the background questionnaire, and 96.2 percent of those completing the background questionnaire completed the exercise booklet. Chapters 5 and 6 document the field operations for the household and prison surveys respectively.

### **1.7 DATA PROCESSING AND MISSING DATA**

After performing quality checks on completed background questionnaires and exercise booklets, field supervisors shipped them to ETS where staff checked the contents of each shipment against the enclosed transmittal form serving as the packing list for the shipment. The background questionnaires were then given to coders who coded the open-ended items, and the exercise booklets were given to readers who scored the open-ended literacy items.

Coding was performed by 20 individuals, 9 working on the background questionnaire and 11 on the exercise booklets, following coding guides developed by scoring supervisors. To check the accuracy of coding in the background questionnaire, items dealing with country of birth, language, wages, and date of birth were checked in 10 percent of the questionnaires by a second coder. In the exercise booklets, 20 percent of all booklets were checked by a second coder who performed a reliability check. The inter-reader reliability for booklets scored by two readers was 97 percent, a number comparing very favorably with the reliability for the 1985 young adult literacy assessment.

The coded responses for the background questionnaire and exercise booklets were then recorded onto scannable answer sheets that were then scanned by ETS staff and transmitted to magnetic tape. The data were then transferred to a database on the main computer for editing and quality control. In a final stage, the data files were examined for nonexistent housing locations, illogical or inconsistent responses,

multiple responses, as well as to insure that the skip patterns had been properly followed and that all data errors had been resolved.

In order to address the issue of missing data, several imputation methods were considered using field test data as well as non-interview report data collected by the interviewers. Three of the five imputation methods made no use of the non-interview report data and the remaining two were informed by the reasons found in the non-interview report. A series of analyses examined the extent to which using each of the five imputation methods affected overall literacy proficiency estimates. Because imputation methods which made no use of the non-interview report data tended to weaken the educational, income, and racial/ethnic differences in literacy scores, they were ruled out, leaving two viable imputation methods. After consulting with others and examining the analyses performed using the two remaining imputation methods, the Technical Review Committee and the Literacy Definition Committee advising the National Adult Literacy Survey project adopted an imputation method for dealing with missing responses. When a respondent failed to answer consecutive assessment tasks and cited a reason related to literacy skills (e.g., “I can’t read these tasks”), the missing tasks were assigned wrong answers. That is, they were scored as if the respondent had attempted and failed the tasks. The extensive processing of the data is detailed in Chapter 7. Chapter 8 provides a discussion of the missing data procedures.

## **1.8 SCALING AND PROFICIENCY ESTIMATES**

The scaling model used for the National Adult Literacy Survey is the three-parameter (3PL) model from item response theory (Birnbaum, 1968; Lord, 1980). This model estimates the probability that an individual will respond correctly to a particular task from a single domain of tasks as a function of a parameter characterizing the proficiency of that individual and three parameters characterizing the properties of a given task in terms of its sensitivity to proficiency, its difficulty, and its non-zero chance of correct response for a multiple-choice task. Item response theory (IRT) models are based on the assumptions of conditional independence (i.e., item response probabilities depend only on a measure of proficiency and the specified item parameters) and unidimensionality (i.e., performance on a set of items is accounted for by a single variable). Thus, a critical part of the data analysis involved the testing of these two assumptions in order to validate the accuracy and integrity of the results.

Because in the National Adult Literacy Survey each respondent was administered relatively few items in a subject area scale, comparing scale scores based on the respondents’ responses to different questions would lead to seriously biased estimates of proficiency. To circumvent this problem, proficiency scores for respondents were estimated using plausible values methodology. Plausible values provide consistent estimates of population characteristics, even though they are not unbiased estimates of the proficiencies of the individuals with whom they are associated. Thus, plausible values are not test scores for individuals in the usual sense. They are merely an intermediate measure used to estimate population

characteristics. Chapter 9 discusses the scaling methodology as well as the calculation of proficiency estimates using plausible values methodology (Mislevy, Beaton, Kaplan, Sheehan, 1993).

### **1.9 ESTABLISHING LITERACY LEVELS**

As previously noted, the results of the National Adult Literacy Survey are reported using three scales: a prose scale, a document scale, and a quantitative scale. The literacy scales, each ranging from 0 to 500, provide a useful way to describe the various types and levels of literacy demonstrated by adults in the population as a whole and in different subpopulations. The scales used an item mapping procedure reflecting response probabilities (RP). Tasks were placed on the scale at the point at which a minimum of 80 percent (i.e., RP80) of respondents at a particular ability level could be expected to complete the task successfully. The scores on each literacy scale represent degrees of proficiency along that particular dimension of literacy. For example, a low score (below 200) on the document scale indicates that an individual has very limited skills in processing information from tables, charts, graphs, maps, and the like (even those that are brief and uncomplicated). On the other hand, a high score (above 375) indicates advanced skills in performing a variety of tasks that involve the use of complex documents.

The literacy scales also make it possible to determine the relative difficulty of the literacy tasks included in the survey. In other words, just as individuals receive scale scores according to their performance in the assessment, the literacy tasks receive different scale values according to their difficulty, as determined by the performance of the adults who participated in the survey. The literacy tasks administered in the National Adult Literacy 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). By examining the tasks within each literacy level, it is possible to identify the types of materials or directives that are more or less difficult for various types of readers. Further, by examining the characteristics of individuals who performed at each literacy level, it is possible to identify factors associated with higher or lower proficiency in reading and using prose, documents, or quantitative materials. Chapter 13 summarizes the establishment of literacy levels for the National Adult Literacy Survey.

Appendices A through R, respectively, contain information about: estimated item parameters, conditioning variables, gamma values, RP80s and item probabilities, the non-interview report form, interviewer's observation guide, English background questionnaire for households, English background questionnaire for prisons, derived variables, codes for continuous variables, birth codes, scoring the variables, sample-specific variables, treatment distribution, estimated composite factors, the code book for windows, and standard errors for Chapters 8 and 10.