

## CHAPTER 8

### MAIN HOUSEHOLD STUDY DATA COLLECTION AND QUALITY CONTROL

Michelle Amsbary, Martha Berlin, Thomas Krenzke, and Leyla Mohadjer, Westat

#### 8.1 INTRODUCTION

To gather information on adults' literacy skills for the National Assessment of Adult Literacy (NAAL), trained staff interviewed a nationally representative sample of about 12,500 adults, aged 16 and older, residing in private households and college dormitories across the United States. Study participants were randomly selected to represent the adult population in the country as a whole. Black and Hispanic adults were sampled at a higher rate than the remainder of the population to ensure reliable estimates of the literacy proficiencies of these groups.

To give states an opportunity to explore the skill levels of their populations, they were invited to participate in a concurrent assessment. Six states elected to participate in the State Assessment of Adult Literacy (SAAL). Approximately 1,000 adults were interviewed in each of the following states: Maryland, Massachusetts, Missouri, New York, and Oklahoma. A larger sample of 1,500 adults participated in Kentucky, because of a request from the state officials that a larger sample be drawn. To allow comparisons of the state and national results, identical instruments were administered to the state and national samples, and the data were gathered at the same time. Unlike the national sample, the SAAL involved no oversampling of Black and Hispanic adults in high-minority areas.

The main household study was conducted from May 2003 through February 2004. Approximately 400 trained interviewers, some of whom were bilingual in English and Spanish, visited households to select and interview adults. Each study participant was asked to spend approximately 90 minutes responding to a series of diverse literacy tasks and answering questions about his or her demographic characteristics, educational background, reading practices, and other areas related to literacy. Also incorporated into the study protocol were a new component to assess oral reading fluency (the Fluency Addition to the NAAL, or FAN) and an alternative assessment to gather as much information as possible about adults with limited English literacy skills (the Adult Literacy Supplemental Assessment, or ALSA).

After completing an interview, the interviewers edited materials for legibility and completeness. The interviewers sent their completed work directly to the data collection contractor's home office, where

the hard-copy assessment booklets and electronic questionnaire data were receipted and processed. Data gathered through computer-assisted personal interviewing (CAPI) were edited by trained staff. Assessment booklets were prepared and sent for scoring, and oral module data files were processed and scored. Final edited and scored data were prepared for analysis.

## **8.2 LISTING**

The implementation of an area probability design such as the one used in the NAAL requires the development of a list of housing units in each second-stage sampling unit, or area segment. This section describes the procedures used to carry out the address listing operation. For the national sample, the NAAL design involved 100 primary sampling units (PSUs) and 1,959 area segments. The design for the state samples involved 74 PSUs (14 of which overlapped with the NAAL PSUs) and 861 area segments (2 of which overlapped with the NAAL segments). Hence, the total household sample was distributed across 160 unique PSUs and 2,818 unique area segments.

During a 12-week period in the summer of 2002, a total of 332,821 housing units were listed in the 2,818 area segments. A field organization of nearly 200 people was assembled to carry out the listing operation.

### **8.2.1 Staff Organization for Listing**

The staff for the NAAL listing operation included 10 regional supervisors and 167 listers. The supervisors reported to one of two field managers, who reported directly to the field director. On average, each supervisor recruited and supervised 17 listers in about 280 segments in his or her region. Listers were recruited in May and June 2002. Of the 167 interviewers hired as listers, 123 had or were currently working for the data collection contractor, 30 had experience with other social research organizations, and 14 were new to social research.

### **8.2.2 Listing Materials**

A segment folder was prepared for each sampled segment. The folder contained (1) a tract map; (2) a segment map; (3) listing sheets on which the lister recorded each address; (4) a segment profile form; and (5) a form containing general comments and any special instructions.

The tract and segment maps in each folder defined and described the sample segments, permitting the listers to identify the exact boundaries of the sampled areas. The size of the segments varied

substantially, depending on the urban or rural character of the area. In dense residential areas, segments may have consisted of one or more blocks. In rural areas, segments may have covered many miles. The 11" × 17" tract map provided an overall picture of the location of the segment within a larger geographic area and within the county. This map gave the listers a geographic context to help them locate the segment. In some segments, the computer program used to generate maps was unable to produce clear and complete maps for the selected geographic areas. In that situation, an enlarged tract map, with clearer boundaries and street names, was also included.

The 11" × 17" segment map was a more detailed picture, showing all streets and other features of the area to be listed. Occasionally, a section of the segment was too dense (i.e., included too many streets), or there was not enough space to print street names and other descriptive information on the map. An enlarged segment map was provided in this circumstance. Listers recorded two things on the segment map: (1) arrows that indicated their route of travel while listing, and (2) the listing sheet line numbers that corresponded to the first and last housing units on each street or boundary.

The U.S. Census Bureau's TIGER System file was used to produce the segment maps. The TIGER file digitized all intersections of geographic boundaries used in the 2000 census. This information was used to generate maps of selected blocks, combinations of blocks, or other geographic units.

The listing sheets were used by the lister to record the complete address, including street name, house number, and, if appropriate, apartment number, of every housing unit encountered. In rural areas, where house numbers were not always available, listers described the housing unit's location in relation to other landmarks.

The Segment Profile Form was used by the lister to collect basic information about the demographic makeup of the segment. The lister also noted the status of new construction or demolition, as well as an estimate of the proportion of the segment used for seasonal dwelling.

The Special Instructions and General Comments Form was used to communicate information to the lister and for the lister to note any special circumstances encountered in the segment.

### **8.2.3 Training Listers**

Of the 167 interviewers recruited for the listing operation, 39 had listing experience within the past 3 years. These listers were trained through home study only, using a listing manual, a listing video, and a home study guide. This guide was a reference manual with practice exercises interspersed

throughout and a final examination. Trainees were instructed to read the manual and complete the exercises and the final examination from the guide. The completed exercises and examination were then mailed to the appropriate supervisor for review. The remainder of the text was used as a reference manual during the listing operation. The experienced listers who completed the home study session were not required to attend the in-person training.

The 10 supervisors and two field managers attended a 1-day in-person training on the listing procedures. The 128 interviewers who had not conducted listing in the past 3 years attended a 2-day, in-person lister training session that included lectures, an audiovisual training presentation, and field practice. Before attending training, these trainees also completed the home study package described above, and brought the completed exercises and final examination with them for the training staff to review and evaluate. The training program covered the fundamental concepts and basic procedures of listing, problematic aspects of listing, special procedures for working in rural areas, and administrative procedures. Listing procedures unique to NAAL were also presented, including instructions for listing group quarters and other structures that did not qualify as housing units, such as military barracks, hospitals, and transient hotels or motel rooms.

A segment near the training site was selected for field practice listing as part of the 2-day training program. The training staff prepared practice segment folders containing the tract and segment maps, as well as listing sheets. Before the training session, the training staff listed all the addresses in the practice segment, noting any problem areas. During training, trainees were taken to the practice segments and required to list the segment individually. Training staff then reviewed the listings with the trainees, checking the recording and discussing any problems.

#### **8.2.4 Listing Operation**

The listing operation began immediately after training and was completed by mid-September 2002. During the initial stages of listing, the lister located the assigned segment using the maps from the segment folder and, when necessary, a local map to verify boundaries. Before beginning to record addresses, the lister “cruised” the segment to verify boundaries, make an approximate count of the housing units in the segment, and correct the segment and tract maps, if necessary.

To keep the listing cost-efficient, very large area segments—those containing 300 or more housing units according to the 2000 census count—were subdivided into smaller, more manageable areas, or chunks, according to instructions provided by statistical staff. One chunk was then selected, with probability proportionate to size, as the area to be completely listed. (Chapter 7 provides more detail on

chunking procedures.) A total of 334 segments had housing unit counts in excess of 300 and were subdivided in this fashion, using a software designed to manage the listing and chunking effort.

If no major problems were encountered during the cruising stage, the lister began the actual address listing operation, starting in the northwest corner of the segment. The starting point and the direction of travel were indicated on the segment map. As the lister traveled through the segment, following the specified listing route, he or she recorded the address of each housing unit on the listing sheet. If no house or apartment number was evident for a housing unit, the lister recorded a detailed description of the unit and its location.

Because the NAAL design was based on the 2000 census data and the listing operation was carried out approximately 2 years after the field operations for the census, relatively few structural changes had occurred in the segments. Hence, in most segments, the difference between the expected and actual numbers of housing units was not great. For the most part, segment boundaries were also still intact and could be easily located.

Field managers, field supervisors, and home office staff monitored the listing effort by using automated progress reports distributed twice a week. Statisticians also monitored the listing operation for yield compared with census figures.

Completed segment listings were returned to the data collection contractor, where they were reviewed for completeness, accuracy, legibility, and adherence to procedures. A segment tracking application was used to track lister assignments and the status of the segment folders. Segment listings were batched for data entry, and address information was then coded, keyed, and entered into the survey control file.

Of the 2,818 segments selected for listing, two segments yielded no housing units. In the remaining 2,816 segments, nearly 332,820 housing units were listed. Of these, approximately 35,500 housing units were selected to form the national sample (approximately 25,500 housing units) and the state samples (approximately 10,000 housing units across the six participating states).

## **8.2.5 Quality Control Procedures**

As described in section 8.2.5.1, quality control checks applied to the listing operation included a thorough review of each lister's initial assignment. Additionally, procedures were implemented to identify

and sample housing units and structures missed during the listing operation. These quality control checks were conducted during the data collection phase of the study and are described in section 8.2.5.2.

### **8.2.5.1 Quality Control of Listing Sheets**

Each lister was required to mail his or her first two completed segment listings to the supervisor for review before working on additional segments. The supervisor reviewed the listings for completeness, accuracy, legibility, and adherence to procedures and provided immediate feedback to the lister. On the basis of this review of the lister's first assignments, the supervisor decided on the type and number of segments to assign to that lister. Trained home office staff conducted a further review of each completed listing.

### **8.2.5.2 Quality Control of the Listing Operation**

As a check on the completeness of the address-listing operation, NAAL interviewers performed two procedures to detect and measure omissions in listing. The hidden housing unit and missed structure procedures were performed during data collection to correct any undercoverage during the listing operation. As the names imply, the procedures separated the detection of missed housing units into two parts—identifying hidden housing units within multiunit structures and detecting completely missed structures or units constructed since the listings were prepared. Section 7.1.3.5 provides the rationale for the hidden housing unit and missed structure procedures, describes how segments and structures were selected for these procedures, and provides the exact number of housing units added through the two procedures. Each of the two procedures is described below.

#### **8.2.5.2.1 Hidden Housing Unit Procedure**

The hidden housing unit procedure was implemented in a sample of the selected housing units. It was designed to detect individual units within listed structures that were not visible to the lister. Such units might be in multiunit structures, such as apartment buildings or duplexes, or they might be separate dwelling quarters within what appeared to be a single-family structure, such as a self-contained in-law apartment in the basement.

The assignment label on the front of the household folder indicated whether the housing unit had been selected for the hidden housing unit procedure. For housing units that required the hidden housing unit procedure, the CAPI screener displayed the text to be read to the respondent to determine whether there were any other living quarters at the address, such as a basement or attic apartment. In multiunit

structures, the interviewer also compared the numbers on mailboxes and doorbells against the listing sheet and looked around the outside of the structure for additional units or entrances, being particularly careful to look for basement, unnumbered, or out-of-the-way apartments that might be hidden and easy to miss.

If no hidden housing units were discovered, the interviewer simply recorded that the procedure had been carried out. If four or fewer additional units were discovered in any housing unit, they were automatically added to the sample, and the interviewer began efforts to conduct interviews in those households. If five or more hidden housing units were discovered in any unit, a subsampling procedure was used to control the number of additional units added to the sample. When this situation occurred, the interviewer called the supervisor or home office for subsampling instructions.

#### **8.2.5.2.2 Missed Structure Procedure**

The missed structure procedure was conducted in a sample of segments. A message on the segment folder instructed the interviewer to perform the procedure before conducting any interviews in that segment. Using the tract and segment maps and the listing sheets, the interviewer recanvassed the entire segment to look for single-family houses or multiunit structures that had been omitted from the listing sheets.

If no missed structures were discovered, the interviewer simply checked a box on the missed structure form to verify that the procedure had been performed. If missed structures were found, the interviewer listed all the newly discovered housing units on the missed structure listing sheet. If five or fewer missed structures were discovered in a segment, they were all automatically added to the sample and the interviewer began efforts to interview in the households. If more than five missed structures were discovered in a segment, interviewers were instructed to call their supervisors to determine whether all or a subsample of units should be added to the sample.

### **8.3 DATA COLLECTION INSTRUMENTS AND INTERVIEWER MATERIALS**

The development and content of the cognitive exercise items, the ALSA, and the oral module—the primary data collection instruments used in the NAAL—are described in detail in chapter 2. What follows is a brief overview of the materials used for the assessments, as well as a discussion of the other materials used during data collection. This includes automated instruments and systems such as the screener, the background questionnaire, the assessment interviewer guide, and the interviewer and supervisor management systems, as well as hard-copy materials such as advance materials, the oral module booklet, noninterview report forms, interviewer manuals, and various field aids. Many of these

materials and instruments were previously developed for the 2001 NAAL field test and revised as appropriate for the main data collection effort.

### **8.3.1 Assessment Materials**

The administration of the core and main assessment, as well as the ALSA, involved the use of many hard-copy materials, in addition to the automated instruments described in the subsequent sections. The development of these materials are discussed in detail in chapter 2, but briefly summarized below.

The data collection effort used 26 unique assessment booklets. Each booklet began with 7 core assessment items, followed by three blocks of main assessment items. Assessment items were arranged into 13 unique blocks and then spiraled to reduce the possibility of question order effects, resulting in 26 booklets.

As in the field test, the administration of the assessment required the use of several stimulus materials, including an almanac, a colon cancer pamphlet, a “Medicare and You” brochure, a newspaper, and a calculator.

The ALSA assessment was produced in a separate booklet and required the use of 9 stimulus materials, including a Coca-Cola can, a box of pancake mix, a yard sale flyer, and a TV guide.

### **8.3.2 CAPI Data Collection Instruments and System Features**

The following sections provide a brief description of the individual automated instruments, the screener, background questionnaire, and interviewer guide, as well as the Interviewer and Supervisor Management Systems.

#### **8.3.2.1 Screener**

The CAPI screener was used to collect household information and to select one or more members of the household for participation in the background questionnaire and the literacy assessment. The screener began with a household enumeration, in which the interviewer recorded the first name of all household members starting with the person (or one of the persons) who owned or rented the home, designated the reference person. The interviewer then entered each household member’s relationship to the reference person, gender, and age. The interviewer asked about race and ethnicity only for household

members aged 16 and older. The screener could be administered to any household member aged 16 or older.

To facilitate the validation of screeners and the subsequent followup of the case, respondents were asked to provide a telephone number where he or she could be reached. Finally, if the housing unit had been selected for the hidden housing unit procedure, the interviewer followed the procedures provided at the end of the screener.

The different selection criteria used for the national and state samples were programmed into the screener instrument. The sampling procedure used to select the appropriate background questionnaire and exercise respondent(s) in each household was implemented by the CAPI system; the interviewer had no discretion about whom to include in the sample. The CAPI screener selected one respondent in households having one to three eligible members and two respondents in households having four or more eligible members.

A Spanish version of the screener was administered by a bilingual interviewer in households where the members spoke only Spanish. Additionally, if the household members did not speak either English or Spanish, interviewers were permitted to use a translator, such as a household member under 16 years of age, a neighbor or friend, or a paid assistant.

### **8.3.2.2 Background Questionnaire**

The background questionnaire was an approximately 25-minute instrument that included questions on a variety of topics, including general and language background; educational background and experience; political and social participation; labor force participation; literacy practices; job training and skills; demographic information; family literacy; household income and welfare participation; health questions; and additional demographics.

Demographic information collected during the screener interview, such as age and gender, was directly imported into the background questionnaire. The CAPI program controlled the background questionnaire instrument flow by using answers to prior questions to determine which questions should be asked and which should be skipped for each respondent. Hard and soft edits were also programmed directly into the CAPI program and inconsistencies were reconciled with the respondent during the interview.

### **8.3.2.3 Interviewer Guide**

Each of the 26 versions of the assessment booklet had a corresponding interviewer guide, which contained instructions for facilitating the interview and guiding the respondent through the assessment booklet and cued the interviewer when to read an instruction or hand stimulus materials to the respondent. The interviewer guide also specified the exact amount and type of assistance the interviewer could provide to the respondent during the assessment. The interviewer guide was programmed as a CAPI application. The CAPI system automatically presented the correct version of the interviewer guide, depending on which assessment booklet had been assigned to the case.

The interviewer guide also contained a module for scoring the respondent's responses to the seven core assessment items. On the basis of the responses to these items, as well as the language in which the core items were administered, the CAPI system implemented an algorithm to determine whether the respondent should proceed with the main assessment or be directed to ALSA. (See section 13.3 for more detail on the core scoring procedure.)

### **8.3.2.4 Interviewer Management System**

An interviewer's assigned cases and other study activities were managed with an integrated software system called the Interviewer Management System. This system was developed for the 2001 field test and revised for the main household study. The Interviewer Management System had various features that were accessed by using one of two laptop modes of operation, stand-alone mode or online mode. The online mode features required access to the data collection contractor's centralized database.

The Interviewer Management System stand-alone mode provided the following capabilities:

- case browse, allowing the interviewer to review assignments;
- hidden housing unit processing;
- status review for the case and its individual tasks;
- the ability to launch and conduct all CAPI instruments;
- creation of zip disks of oral module recordings; and
- entry of status codes and other information on an Electronic Record of Calls (EROC).

The Interviewer Management System online mode provided the following capabilities:

- data transmission;
- time and expense reporting;
- shipment of case materials; and
- e-mail.

The five primary functions of the Interviewer Management System were (1) Browse Case, (2) Browse Person, (3) Data Transmission, (4) E-mail, and (5) Ship Disk. Each function is briefly discussed below.

### ***Browse Case***

The Browse Case window displayed the following information about each of the interviewer's assigned cases:

- case identification (ID) number;
- street address, city, ZIP code, and state;
- overall interview status and status date; and
- assigned assessment booklet type for sample person #1.

The Browse Case window also contained the Activity Log. The Activity Log displayed a history of previous EROC entries. The items included in this display were the following:

- the contact date;
- the contact status or result;
- the contact type (in-person, telephone, or mail);
- who was contacted;
- comments about the contact; and
- the appointment date, if applicable.

The EROC allowed the interviewer to enter interim status codes only. Finalized statuses were entered by the supervisor, using the Supervisor Management System application (see section 8.3.2.5), or by the CAPI system automatically upon completion of each interview task.

The Browse Case window also provided the functionality for updating addresses and completing the hidden housing unit procedure. If an address correction was needed, as determined during the study introduction that preceded the screener, the interviewer entered revised address information into the Interviewer Management System. This action provided information for updating the home office files to be used in conducting validation interviews or sending refusal conversion letters. The hidden housing unit button was available only for the cases that had been selected for the procedure. Interviewers entered the number of housing units to be added, as well as the street address for these cases. The Interviewer Management System then added the new cases to the interviewer's Browse Case assignment.

### ***Browse Person***

When a sample person was identified during screening, sample person-level data were added to the system. For completed screeners in which one or two sample persons were selected, Browse Person was used to view sample-person-level information, such as sample person ID number, name, age, gender, telephone number, sample-person-level status, status date, and assigned assessment booklet type. Browse Person functionality was also used to complete the sample person-level tasks, such as the background questionnaire, the assessment, and the oral module. The Interviewer Management System controlled the order of administration of the instruments by requiring that the background questionnaire be finalized before the interviewer guide associated with the assessment became available.

### ***Data Transmission***

The Data Transmit button was used by the interviewers to enter their time and expense data, transmit case data, and record shipments of cases to the data collection contractor. The data transmission and shipment functions are discussed in more detail below.

Data transmission involved sending electronic information from the interviewer's laptop to the home office and sending new study data to the interviewer's laptop. Items that were sent from the home office to the interviewer's laptop included new or transferred assignments from the supervisor and case status code updates entered by the supervisor. Items that were sent from the interviewer's laptop to the home office included EROC entries, interview data for completed or partially worked cases, and data concerning finalized cases that were ready to be shipped to the home office.

As part of the preparation for shipping finalized case materials, interviewers indicated in the Interviewer Management System which cases would be included in each shipment, as well as the tracking number used for the mailing.

### *E-mail*

Outlook Express was used to send and receive e-mail. Interviewers were in frequent e-mail communication with their supervisors and other interviewers.

### *Ship Disk*

The Ship Disk process transferred the voice recordings collected during the oral module from the laptop to a zip disk. Interviewers were given a zip drive to assist in this process.

### **8.3.2.5 Supervisor Management System**

The Supervisor Management System, a component of the Data Management System, was designed for the 2001 field test and revised as necessary for the main study. Supervisors used the Supervisor Management System to manage the case work within their region. The numerous functions of the Supervisor Management System are described below:

- case review, assignment, reassignment, and unassignment to interviewers;
- case search using the following criteria: PSU, case ID, interviewer, status, and status date;
- assignment of the final status of cases, at both the household and sample person levels;
- review of time and expense data that were recorded by interviewers;
- report production (see section 8.7.2. for a detailed discussion of the reports); and
- tracking of the distribution and use of the assessment and ALSA booklets.

### **8.3.3 Interviewer Materials**

The administration of the NAAL interview required numerous hard-copy interviewer materials, including several advance materials, the household folder, handcards, noninterview report forms, interviewer manuals, and the oral module booklet. These materials are discussed in the following sections.

### 8.3.3.1 Advance Materials

During preparation for the 2001 field test, considerable effort was made to develop introductory materials that would convince respondents of the study's legitimacy and importance. These materials were revised slightly for the main study.

**Advance letter.** Before the interviewer's first contact with the household, the home office mailed an advance letter with a brochure. The letter introduced the study, identified the sponsor, stated the study's purpose, and asked for cooperation. It also provided the toll-free study hotline number, as well as the name and telephone number of the National Center for Education Statistics (NCES) Project Officer.

**Brochure.** The informative and attractive brochure included with the letter explained the study in detail and emphasized the importance of participation.

**U.S. Department of Education letter of introduction.** This letter was signed by the NCES Project Officer and verified that the interviewer was an authorized representative of the U.S. Department of Education. The letter was used whenever respondents needed further convincing that the study and interviewer were legitimate.

**Community authorization letter.** This general letter was intended to be displayed to apartment managers, postal employees, police departments, or other professional people whom interviewers might encounter in the community. It provided assurances that the interviewer was not selling or soliciting but was a trained professional working on a government-sponsored education study.

**Sorry-I- missed-you card.** This card was left when the interviewer visited a household and no one was home. Interviewers often personalized the card with a brief message or left their name and telephone number on the card; this provided some familiarity and recognition when the interviewer returned to the household.

**Refusal conversion letters.** Several versions of refusal conversion letters were developed and sent to households that refused to participate in the study. See section 8.8 for a list of these letters.

Spanish-language versions of the advance letter, brochure, U.S. Department of Education letter of introduction, and sorry-I-missed-you-card were produced and distributed as well.

### **8.3.3.2 Household Folder**

One household call record folder was produced for each sampled housing unit in the study. The household folder helped the interviewers keep track of the status of all cases in their assignment.

A label on the cover specified the case ID, the address of the housing unit, the listing sheet line number associated with the housing unit, the barcode number of the assessment booklet preassigned to the case, and the case control code. It also indicated whether the housing unit had been selected for the hidden procedure.

The front of the folder also contained a study introduction, for convenient access at the doorstep, and included an address verification question. The Spanish translation of the study introduction was located on the inside front cover. Step-by-step instructions for performing the hidden housing unit procedure were also included on the inside front cover, as well as space for recording any necessary notes.

The Record of Actions was located on the back cover of the household folder. Interviewers used the Record of Actions grid to record the status and outcome of every contact attempt with the household. For each contact, the interviewer recorded the following information:

- date, day of week, and time;
- contact type (in-person, telephone, or mail);
- result code of appropriate instrument, for both sample person #1 and sample person #2;  
and
- any applicable comments.

The interviewer stored the case-specific assessment booklet in the household folder, as well as any noninterview report forms, oral module interviewer guides, ALSA booklets, or assessment booklets used with a second sample person, if required for that case. This entire package was returned to the home office when the case was completed. If a case was transferred to another interviewer, the household folder was transferred as well.

### **8.3.3.3 Handcards**

A bound handcard booklet was designed to facilitate the flow of the screener and background questionnaire interviews and improve the efficiency of the sample person's reporting because the response options would not have to be read aloud for every question. Two handcards listing race and ethnicity response options were used for the screener. The background questionnaire used 10 handcards listing the sets of response categories used most frequently throughout the interview. The handcards were translated into Spanish, with the translation bound to the flip side of the handcard booklet. Handcard instructions, indicating which handcard was to be used for each item, appeared at the top of the applicable CAPI screen.

### **8.3.3.4 Noninterview Report Forms**

When a household member did not complete the screener, or when a sampled respondent did not complete the background questionnaire, the assessment booklet or ALSA, or the oral module, the interviewer was required to complete a noninterview report form. One version of the form was used to document nonresponse at the screener level; another version was used for nonresponse resulting at the background questionnaire, assessment, ALSA, or oral module level. The information collected on these forms served two important purposes: (1) field supervisors reviewed the forms to determine the case's potential for nonresponse conversion and (2) the data collected on the form were processed for nonresponse analysis.

The screener noninterview report form was completed if the sampled address was determined to be vacant or not a housing unit, or if the interviewer was unable to complete a screener at that address. In the latter case, the interviewer provided information about attempts to contact the household and the reason for noncompletion. If a household member refused to participate, the interviewer described the reasons, in the respondent's own words, as well as the perceived strength of the refusal. The interviewer also provided any other information that might help another interviewer contact the household, complete the screener, or both, such as what language was spoken in the household and whether a refusal letter would be helpful.

The sample person noninterview report form collected similar information for the background questionnaire, the ALSA, the assessment, and the oral module, including the specific reasons the respondent did not complete the particular instrument(s) and additional information, such as the type of disability, the non-English language spoken, details about reading and writing difficulty, or circumstances surrounding the refusal.

### **8.3.3.5 Interviewer Manual**

The study-specific interviewer manual included an introduction to the study and an overview of interviewer responsibilities. The text covered field materials and procedures for locating sampled households; contacting respondents; and administering the screener, the background questionnaire, the assessment booklet, the ALSA, and the oral module. Question-by-question specifications for the screener, the background questionnaire, the interviewer guide, and noninterview report forms were included as well. The interviewer manual also contained information on maintaining quality control procedures, using the Interviewer Management System features, keeping records, completing the Time and Expense Report, and reporting to the supervisor. A detailed table of contents and section markers helped the interviewer locate specific information in the manual. Interviewers received the manual at training.

### **8.3.3.6 Oral Module Booklet**

The oral module booklet contained the tasks that the respondent was required to read aloud during the oral module component of the interview. As described in section 5.3.4.4, respondents wore a headset with a microphone that recorded samples of their reading and answers to questions directly onto the interviewer's computer. The oral module booklet contained eight unique reading passages, spiraled to create 16 different sets consisting of two passages each. The CAPI oral module instrument instructed the interviewer which set of passages to administer to each respondent.

The reading passages were followed by one list of digits, one list of letters, one list of English words (subdivided into three sections that got progressively harder), and one list of English pseudowords (subdivided into three sections that got progressively harder). These were presented in the same order for all respondents. (See section X.X for a detailed description of the tasks included in the Oral Module Booklet.)

## **8.4 FIELD STAFF TRAINING**

The following sections describe the training of the field supervisors and interviewers. The training plan adopted included one large session, with the interviewers divided into 20 small groups according to supervisory region. The field supervisor served as the assistant trainer in each room and was paired with a lead trainer. NCES and various contractor and subcontractor staff members monitored the sessions.

A challenge of the training plan was to prepare field staff to both conduct traditional interviews and administer literacy assessments. Interviewers were trained to take an active role in conducting the screener and the background questionnaire and to be prepared to answer any questions the respondent might raise. In the role of assessment administrator, however, the interviewers had to remain very much in the background, observing and facilitating but intervening only at certain well-defined points and refraining from offering help in completing the literacy tasks even if it was requested.

#### **8.4.1 Approach to Training**

The basic approach to interviewer training was to maximize trainees' involvement and participation in the training, to provide ample opportunity for supervisory staff to observe and evaluate trainee performance, and to provide trainees with detailed reference documents.

Each training room had a lead trainer and an assistant trainer responsible for approximately 15 to 18 trainees. The training staff was composed of members of the data collection contractor's staff. Lead trainers consisted of home office staff, the four NAAL field managers, other employees with training experience, and supervisors from other data collection projects. The 20 regional supervisors served as the assistant trainers.

#### **8.4.2 Train-the-Trainer Trainings**

Approximately 50 lead and assistant trainers were trained on the NAAL training program during a training session in late March 2003. The 4-day training was a simulation of the interviewer training program (described in detail in section 8.4.5), although the pace was accelerated because of the experience level of the group. This simulation of the interviewer training program not only prepared the supervisors for their subsequent responsibilities but also provided a dress rehearsal for training staff and an opportunity to evaluate and refine the training materials. After completing the training, lead and assistant trainers were prepared to lead small groups of interviewers through scripted, interactive reviews of the NAAL data collection instruments.

Following the lead and assistant trainer training, a separate training session was conducted in April 2003 for the staff who would be supporting each training room's software and hardware needs, referred to as runners and data display operators. Given the reliance on CAPI instruments throughout the interviewer training, it was necessary to train more than 50 people on how to navigate and resolve problems with the NAAL CAPI applications. An abbreviated simulation of the interviewer training was

conducted over a 2-day session, focusing exclusively on the training sessions that involved CAPI applications and the use of the computer.

### **8.4.3 Supervisor Training**

A 1½-day training was held for the field managers and supervisors immediately following the train-the-trainers training session. Supervisor training was conducted by the NAAL field director, with support from NAAL systems staff on the technical aspects of the Supervisor Management System. The training covered management techniques and duties specific to the NAAL data collection. The supervisor manual was distributed at the training.

After completing their training, field managers and supervisors returned home to prepare to train interviewers and assume supervisory responsibility immediately after interviewer training.

### **8.4.4 General Interviewer Techniques**

Novice interviewers received 4 hours of in-person training on General Interviewer Techniques (GIT) prior to project-specific training. The in-person GIT training program included an audiovisual presentation, interactive participation, written exercises, and a question-and-answer period. The training introduced the interviewers to survey research; provided examples of survey questions, recording conventions, and interviewing terminology; and taught them basic listening and probing skills for obtaining accurate data. The interviewers completed exercises on applying probing techniques and answering respondent questions. The importance of data quality was also reviewed. This training was in addition to the GIT home study and exercises, required of all interviewers.

CAPI Train, a self-administered tutorial that introduced the procedures for conducting a CAPI interview, was completed by most interviewers, as part of their home study package before training. The tutorial instructed trainees on types of questions, function keys, and special commands. The training also included practice in logging on to the computer and using the keyboard, particularly the function keys used to manage the flow of the instruments.

For the handful of interviewers who were hired late in the recruitment process and did not receive their laptops prior to training, a separate 2-hour CAPI Train session was held after GIT training. Interviewers completed the self-guided tutorial at their own pace. The session was supervised by NAAL systems staff.

## **8.4.5 Project-Specific Training**

NAAL project-specific training for the 342 interviewers consisted of a 6-day, in-person training program, preceded by home study. The training was conducted in late April 2003. Twenty groups of interviewers were trained concurrently. Supervisors (assistant trainers) were assigned to the training room with the interviewers from their region. Holding numerous simultaneous training sessions at one site allowed the NAAL field director, data collection contractor home office staff, and NCES staff to observe all training sessions while maintaining a manageable number of interviewers in each training room.

Because of interviewer attrition, 59 interviewers from across the country were recruited and trained at a session held in August 2003. Two additional small attrition trainings for a total of 10 interviewers were held in fall 2003. These training programs, led by NAAL project staff, were identical to the program used at the initial interviewer training session.

### **8.4.5.1 Interviewer Training Materials**

The training materials were carefully scripted to cover every concept that the interviewers needed to know, and the scripts were organized into training guides. The elaborate preparation of training materials accomplished two purposes. First, it achieved standardization, which is particularly important when a large staff of interviewers is being trained in separate sessions. Second, it allowed all trainers to study the training guides, rehearse their roles, and be completely prepared for training. This was particularly important in a training effort that required a large training staff. The scripted materials eliminated the necessity for the trainer to improvise. This preparation allowed the NAAL training sessions to move smoothly and on schedule, which gave the interviewers the confidence that they were being trained by knowledgeable people.

### **8.4.5.2 Interviewer Training Techniques**

The general approach to training centered on five basic training techniques that have been extensively used and refined by survey operations professionals over the past 30 years. The following paragraphs briefly describe the five techniques and how they were used for training on NAAL.

**Home study.** About 2 weeks before training, interviewers received a home study package and their laptop computer. The home study included an overview of NAAL, instructions on how to set up and test the laptop computers in the interviewers' homes, an e-mail tutorial, the CAPI Train tutorial and

exercises, and the GIT home study guide and exercises. Completed exercises were collected and reviewed by training staff at the in-person training.

**Demonstration.** The first session at training was a videotaped demonstration of the entire NAAL interview. The demonstration interview introduced the trainees to the NAAL instruments and gave them an idea of their role and responsibilities on the project. The trainees were able to see the overall flow of the interview before receiving instruction on each instrument.

**Interactive lecture.** This technique provided trainees with detailed instructions for administering the questionnaires. The lead trainer used a scripted lecture to present the basic concepts of the instrument to the entire group of trainees. Trainees took turns playing the role of interviewer and asking the questions, while the lead trainer provided responses from the script. The lead trainer's script included instructions to interrupt the script at appropriate times to review certain sections of the interviewer manual, point out some of the less obvious features of the instrument, or explain certain terms. All trainees were required to follow along on their computers and enter the responses provided by the trainer. Several runners assigned to the room ensured that trainees entered the correct responses and were all on the correct item. A response was entered into a laptop by a trained data display operator, and then projected on a screen in front of the group. Trainees were instructed to check their entry against the entry on the screen. Interactive lectures were used for the initial presentations of the screener, background questionnaire, interviewer guide, ALSA, and oral module. The scripts used for the interactive lectures presented increasingly complex scenarios so that trainees became familiar with the various types of cases they would encounter.

**Practice exercises.** Written exercises reinforced and tested trainees' comprehension of certain concepts. They were particularly well-suited for evaluating the trainees' comprehension of some of the more complicated instrument issues, such as navigating the screener enumeration grid, scoring the core exercise items, and collecting industry and occupation information in the background questionnaire.

**Dyad role playing.** Role playing provided additional practice and gave trainees a feeling for the overall flow of the interview. Trainees were arranged in pairs (dyads), as designated by the training team. One member of each pair was given a scripted copy of the interview instruments, complete with data entry instructions, and played the role of the respondent while the other trainee conducted the interview (played the role of the interviewer). With the next script, the members of the pair reversed roles. Two role playing scripts were used. The scripts began with the screener and ended with the administration of the oral module.

Additionally, paid respondents recruited by a local focus group facility were brought in toward the end of the training session. Interviewing these respondents gave the trainees the opportunity to conduct a real nonscripted interview before working their first case assignment. Training staff observed these practice interviews and provided feedback after the interview.

### **8.4.5.3 In-Person Training Program for Interviewers**

Most of the 6-day interviewer training was devoted to teaching procedures for administering the data collection instruments—screeners, background questionnaire, assessment booklet and interviewer guide, ALSA, and oral module. In addition, instruction was provided on gaining respondent cooperation, locating households, using the Interviewer Management System, assigning status codes, and completing administrative forms. Table 8-1 presents an overview of the training program.

Training interviewers to administer the exercises presented a particular challenge. The role of the assessment administrator was different in important ways from that of an interviewer, requiring interviewers to switch roles in the middle of the interview. During the administration of the screener and background questionnaire, a dynamic interaction took place between the interviewer and the respondent. Although the interviewer needed to remain neutral and avoid leading the respondent, he or she provided reassurance and encouragement. The administration of the assessment exercises, in contrast, required the interviewer to take a much more passive role, observing the respondent's performance without intervening (except as directed in the interviewer guide) and studiously avoiding any temptation to provide assistance, even when help was requested.

**Table 8-1. Overview of the NAAL study-specific interviewer training session: 2003**

| Day | Topic   | Presentation mode                              |
|-----|---|--|
| 1   | Introduction and overview of the study  | Plenary session                                |
|     | Demonstration of the NAAL interview   | Plenary session                                |
|     | Advance materials   | Learning community                             |
|     | Procedures for gaining respondent cooperation                                 | Learning community                             |
|     | Locating households   | Learning community                             |
|     | Use of Interviewer Management System  | Learning community                             |
| 2   | Screener  | Interactive lectures                           |
|     | Screener enumeration grid practice  | Individual exercise                            |
|     | Background questionnaire  | Interactive lectures                           |
| 3   | Background questionnaire  | Interactive lectures and individual exercises  |
|     | Assessment booklet and interviewer guide                                      | Interactive lectures                           |
|     | Core scoring procedures   | Interactive lectures and individual exercises  |
|     | Administrative procedures   | Interactive lectures                           |
| 4   | Oral module   | Interactive lectures                           |
|     | Screener, background questionnaire, core and main assessment, and oral module | Dyad role playing                              |
| 5   | Day in an interviewer's life  | Interactive lectures                           |
|     | Gaining respondent cooperation  | Round table discussion                         |
|     | Live respondent practice  | One-on-one                                     |
|     | Meeting with supervisors  | Learning community                             |
| 6   | ALSA  | Video, interactive lectures, dyad role playing |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

The issues involved in making the switch from interviewer to assessment administrator were discussed in a lecture conducted by a lead trainer. In addition to instructing interviewers on the mechanics of administering the assessment, the trainer spent time discussing problems that might occur in the administration of the assessment, such as those arising for respondents with limited English-reading abilities, respondents with physical or mental conditions that might affect their performance on the assessment, and the special needs of the elderly population. The trainees then practiced administering the assessment, received instruction on core scoring procedures, and reviewed recordkeeping procedures specific to the assessment booklets.

In general, interviewers were discouraged from making assumptions about a respondent's capacity or ability to complete a study instrument; they were instructed to allow the respondent to attempt the questionnaire and the assessments as long as the respondent was willing. In addition, the training materials emphasized the fact that disabilities that might prevent the respondent from completing one instrument might not prevent him or her from completing another instrument. For example, although a

blind respondent would be unable to complete the assessment booklet, he or she would not be prevented from responding to the background questionnaire. Conversely, although a respondent who had insufficient English skills might be prevented from responding to the background questionnaire, he or she might be able to attempt the assessment booklet. Although the CAPI system required that the background questionnaire be completed before the assessment was administered, a technique was developed for situations such as these, to allow interviewers to complete the background questionnaire and attempt the assessment.

During training, evening meetings were held with all lead and assistant trainers, other data collection contractor home office staff, and field management staff to discuss any problems that had arisen during the day. Minor modifications to the training program or schedule were discussed as a group. Any necessary changes or clarifications to the materials were made and distributed to the training staff.

Trainees with potential performance problems were identified and remedial measures were discussed at these nightly meetings. Such trainees were closely observed and were paired during dyad role plays with a staff member who could assist them during the mock interview. They were also required to attend evening practice sessions at which they could focus on the element of the interview causing the most problems, such as navigating the screener enumeration grid or using the Interviewer Management System. One-on-one conversations were held with these trainees about their progress. At the end of training, eight trainees who were unable to master the procedures and techniques required for the job were released from the study.

The trainee group included 60 Spanish-speaking interviewers. Following the study-specific training, these interviewers were assembled into three training communities, which were led by a Spanish-speaking field manager, supervisor, and home office staff member. During this 6-hour training session, interviewers worked with the Spanish translation of the screener, the background questionnaire, the core assessment items, the oral module, the ALSA, and advance materials. This gave the supervisors the opportunity to assess the Spanish-speaking abilities of the bilingual interviewers.

## **8.5 CONDUCT OF THE FIELD WORK**

The NAAL field period began in May 2003, immediately following the completion of the interviewer training session, and lasted for approximately 40 weeks, until the beginning of February 2004. The national and state samples were worked simultaneously.

The following sections describe the field operations, including the general approach, the schedule and production, and the reporting systems used to manage the field effort.

### **8.5.1 Field Organization**

The national and state study components were carried out by a large field organization, headed by the NAAL field director, who reported directly to the data collection contractor project director and was supported by 4 field managers and 20 field supervisors across the United States. The supervisors oversaw an interviewing staff of approximately 350 interviewers. This section presents a general description of the field organization and the responsibilities of the staff at each level.

#### **8.5.1.1 Recruiting Field Staff**

Field staff were recruited and hired directly, not through interviewing services. Interviewers were hired from the areas in which the interviewing assignments were located. The primary source of potential field staff was the data collection contractor's computerized field personnel file, containing information on more than 12,000 persons who have worked on their field studies in the previous 4 to 5 years. The file excludes those who received an unsatisfactory review during a previous assignment. This system can produce lists, by geographic area, of available field personnel who meet the qualifications for a project. The system contains demographic information on languages spoken, special field skills, and time and geographic availability. Project evaluations are also in the system, including productivity, accuracy, cooperation, dependability, and length of service for each project. The four field managers and 20 regional supervisors were hired in late 2002 and early 2003.

When recruiting interviewers, supervisors and field managers assessed both the basic skills and personal traits of applicants. An interviewer had to have basic reading and computational skills and be able to follow instructions. Computer skills were desirable as well. Desirable personality traits included receptivity to others' ideas, open-mindedness, and motivation. Additionally, a respondent's willingness to grant an interview often depends on his or her initial perception of the interviewer. Candidates who would appear "neutral" to the target population were favored.

Of the 342 interviewers recruited, 334 successfully completed training and began work on the study. The field organization included approximately 60 interviewers who were bilingual in Spanish and English.

### **8.5.1.2 Field Staff Attrition and Replacement**

The level of interviewer attrition was greater than that experienced for the 1992 NALS, but was consistent with that found on recent field studies of similar size and length of field period. The leading causes of attrition were poor performance, such as inadequate interviewer production rates or a high number of hours per completed case; personal or family illness; and acceptance of a full-time position elsewhere. A total of 86 interviewers were released throughout the field period, and another 110 resigned, for a total attrition rate of 47 percent. Different approaches were used to deal with attrition problems depending on when and where they occurred. In some cases, new interviewers were hired and trained. In other cases, other interviewers working in the PSU were able to complete the remaining work, or interviewers from other areas traveled to the PSUs where the attrition occurred.

To compensate for attrition and slow production in a few areas, an additional 63 interviewers were recruited in July 2003; of these, 59 completed training. Two small-scale attrition trainings were conducted in September and November 2003 as well.

The characteristics of field workers, including all recruited interviewers, supervisors, field managers, and the field director, are shown in table 8-2. Nearly half (224, or 46 percent) had worked previously for the data collection contractor, and a similar number (216, or 45 percent) had worked as interviewers for other field organizations. The field staff was primarily middle aged, with most (322, or 67 percent) between the ages of 30 and 59, only a small number (13, or 3 percent) under 30, and slightly fewer than one-third of them (148, or 31 percent) aged 60 or older. Like most field staffs, the majority (380, or 79 percent) were female.

**Table 8-2. Characteristics of field workers on NAAL main study, by gender, age, experience, and assignments: 2003**

| Characteristics  | Total persons | Female  | Male    |
|--|---------------|---------|---------|
| Total persons  | 483           | 380     | 103     |
|  | Percent       | Percent | Percent |
| Age (years)  |               |         |         |
| Total  | 100.0         | 100.0   | 100.0   |
| 20–29  | 2.7           | 2.6     | 2.9     |
| 30–39  | 11.0          | 11.1    | 10.7    |
| 40–49  | 22.3          | 22.1    | 23.3    |
| 50–59  | 34.5          | 29.1    | 33.3    |
| 60–69  | 22.4          | 25.2    | 23.0    |
| 70–79  | 6.8           | 7.8     | 7.0     |
| 80+  | 0.5           | 1.0     | 0.6     |
| Survey research experience with the data collection contractor (years since hired) |               |         |         |
| Total  | 100.0         | 100.0   | 100.0   |
| ≤ 12 months  | 45.0          | 51.4    | 46.4    |
| 1–4 years  | 34.5          | 40.0    | 34.6    |
| 5–9 years  | 11.1          | 4.9     | 9.7     |
| 10–14 years  | 6.1           | 7.8     | 6.4     |
| 15–19 years  | 2.9           | 1.0     | 2.5     |
| 20–24 years  | 0.5           | 0       | 0.4     |
| Field assignments worked   |               |         |         |
| Total  | 100.0         | 100.0   | 100.0   |
| 1–4  | 80.5          | 83.5    | 81.1    |
| 5–9  | 13.7          | 11.7    | 13.3    |
| 10–14  | 2.6           | 3.9     | 2.9     |
| 15–19  | 1.6           | 0       | 1.2     |
| 20–24  | 0.5           | 1.0     | 0.6     |
| 25–29  | 0.3           | 0       | 0.2     |
| 30+  | 0.8           | 0       | 0.6     |

See notes at end of table.

**Table 8-2. Characteristics of field workers on NAAL main study, by gender, age, experience, and assignments: 2003—Continued**

| Characteristic                              | Total percent | Female<br>(percent) | Male<br>(percent) |
|---|---------------|---------------------|-------------------|
| Survey research experience: other companies |               |                     |                   |
| Total                                       | 100.0         | 100.0               | 100.0             |
| None  | 41.1          | 58.3                | 44.7              |
| 1–4 years                                   | 39.7          | 31.1                | 37.9              |
| 5–9 years                                   | 8.4           | 5.8                 | 7.9               |
| 10–14 years                                 | 4.2           | 1.9                 | 3.7               |
| 15–19 years                                 | 3.4           | 1.0                 | 2.9               |
| 20–24 years                                 | 1.6           | 0                   | 1.2               |
| 25+ years                                   | 1.2           | 1.9                 | 1.7               |
| Highest level of education                  |               |                     |                   |
| Total                                       | 100.0         | 100.0               | 100.0             |
| Unknown/no data                             | 0.3           | 0                   | 0.2               |
| Some high school                            | 0             | 1.9                 | 0.4               |
| High school                                 | 15.3          | 3.9                 | 12.8              |
| Vocational certificate                      | 7.9           | 10.7                | 8.1               |
| Some college                                | 40.5          | 31.1                | 36.8              |
| Bachelor's degree                           | 21.0          | 20.4                | 20.0              |
| Graduate work                               | 3.4           | 9.7                 | 14.5              |
| Graduate degree                             | 11.6          | 22.3                | 13.2              |

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

## 8.5.2 Field Management

The management of the NAAL data collection effort included the project director, the field director, field managers, and regional supervisors. The following sections describe the reporting structure of this organization, as well as the procedures and tools used to assist in the reporting.

### 8.5.2.1 Reporting Structure

The data collection contractor home office staff who oversaw the NAAL field organization included the project director and several supporting staff members. The field director and regional supervisors were located in the field. The field director coordinated all activities related to field operations and kept in close touch with the four field managers to address issues of production, cost, response rates, shipment of closed-out work, and other issues.

For purposes of field operations, the 160 NAAL PSUs were divided into 20 regions, each headed by a supervisor who typically lived in the region. Each field manager had responsibility for five regions and five regional supervisors. The field supervisor's primary responsibility was overseeing the work of an average of 15 to 18 interviewers in his or her region.

An important part of the supervisor's job was determining the optimal flow of work to each interviewer. On the basis of the weekly conference call, the supervisor decided when the interviewer was ready for an additional assignment. Supervisors tried to maintain a balance between somewhat competing goals—keeping interviewers supplied with enough work to stay productive and not allowing cases to languish by giving an interviewer more work than he or she could close out in 2 or 3 weeks.

### **8.5.2.2 Reporting Procedures and Tools**

The smooth progress of field work depended on the ongoing monitoring of the interviewers' work and regular communication among all members of the NAAL project staff and NCES. The following sections describe the major mechanisms and procedures used for reporting during the NAAL field period.

Interviewers were required to contact their supervisors by telephone at a regularly scheduled time once a week to discuss all aspects of their work (response rates, production and cost performance, and quality control results). Each outstanding case in the interviewer's assignment was reviewed and discussed. The supervisor and interviewer discussed any problems reflected in the Supervisor Management System-generated data collection reports, such as low response rates or a high number of hours per complete case. (See section 8.7.2 for a discussion of these Supervisor Management System reports.)

At least once a week, each supervisor had a telephone conference with his or her field manager to discuss progress in the region. Discussion centered on the week's Supervisor Management System reports as well as on current progress as reported to the supervisor during the interviewers' weekly calls. The weekly conferences between field managers and supervisors were used to discuss problems in the region, the prospects and plans for completing the remaining work, and what help, if any, the supervisor needed in order to complete all work in the region by the end of the field period. The results of quality control procedures were also discussed. If the quality control reports indicated problems with the quality of an interviewer's work, appropriate steps to correct the problem were discussed.

Additionally, the field director conducted a formal weekly telephone call with the field managers. The discussion began with a review of the Supervisor Management System reports. These reports were

used during the weekly calls to identify any problems in an area, discuss plans for completing the remaining work, and determine whether any help was needed to meet production goals.

Once a week, a home office staff meeting was held with the project director, the field director, and other home office staff members. The results for each region were reviewed, and any studywide problems were reviewed (e.g., interviewer attrition, Interviewer Management System software or hardware issues, or distribution of supplies and materials). Strategies for solving problems were discussed and passed on to the field managers and other staff for implementation.

Twice a month, a summary of data collection progress was sent to the NCES Project Officer. Finally, once a month, NCES and the NAAL contractors attended a meeting at which field progress and problems uncovered during the review of work were discussed. Any important changes in the field work strategy were discussed before implementation.

## **8.6 DATA COLLECTION OPERATIONS**

The following sections describe the general approach to the NAAL data collection operation, the specific schedule and plan for production, and the procedure used to effectively administer the interview.

### **8.6.1 General Approach**

The NAAL field effort used an approach that has been effective for many previous surveys involving large, complex, in-person data collection operations, including the 1992 National Adult Literacy Survey. Under this approach, the field effort occurs in three overlapping stages:

- Initial phase. Each area segment<sup>1</sup> is assigned by the regional supervisor to an interviewer, who follows certain rules in making a prescribed number of calls to every sampled housing unit in the segment.
- Reassignment phase. Cases that do not result in completed interviews during the initial phase are reviewed by the regional supervisor, and a subset is selected for reassignment to another interviewer in the same or a nearby PSU.
- Special nonresponse conversion phase. The home office assembles a special traveling team of the most experienced or productive interviewers to perform a nonresponse conversion effort, under the supervision of a subset of the field supervisors.

---

<sup>1</sup> Census blocks or groups of blocks within PSUs.

The assignments in the initial phase are controlled by the regional field supervisor. In NAAL, the supervisors had two local interviewers available in most PSUs. Each area segment was assigned to one of the interviewers on the basis of the racial/ethnic composition of the area and the proximity of the segment to the interviewer's home.

During the initial phase, interviewers were instructed to make up to four in-person calls to the household to complete a screener and up to four additional in-person contacts—after completing the screener—to administer the background questionnaire, the assessment, and the oral module, which had to be completed during the same visit. After the prescribed number of in-person attempts to complete the instrument without contacting a respondent, the interviewer consulted with the supervisor to determine further attempt and contact strategies.

To maximize the chances of finding respondents at home, most contacts were made during prime interviewing hours (3 p.m. to 9 p.m. on weekdays and 10 a.m. to 9 p.m. on Saturdays and Sundays). Contacts at each housing unit were to be staggered on different days of the week and at different times of the day. All calls to complete the screener had to be made in person. If the screener was completed and the background questionnaire, the assessment, and the oral module could not be completed in the same visit, the interviewer was permitted to use the telephone to set an appointment to administer these instruments in person. The initial phase was considered complete when the interviewer reported a definitive outcome for the case or when the full complement of calls had been made.

Interviewers mailed completed cases to the home office twice a week. Copies of the entire segment folder and its associated materials were returned when the interviewer completed all possible cases in the segment. At the time the assignment was made, the interviewer was given 2 to 3 weeks (depending on the size of the segment) to complete the initial effort for all cases in the segment. Most productive interviewers were able to handle up to five segments simultaneously during the initial phase. Less productive interviewers were given only two or three segments at a time.

### **8.6.2 Schedule and Production**

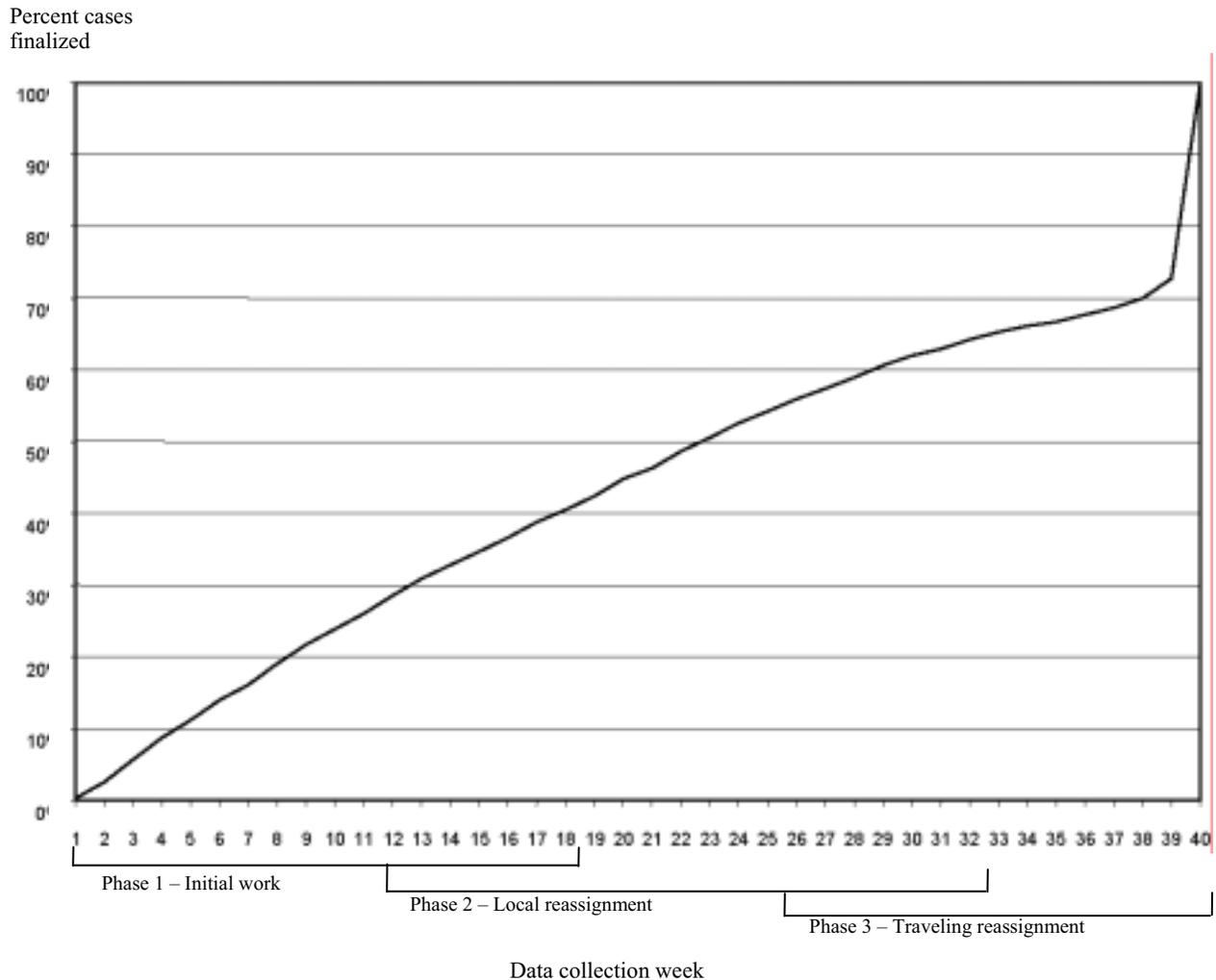
The original plan for the NAAL field effort envisioned a 26-week (6-month) field period, in which 4 months would be used to complete the initial complement of calls to all assigned households and 2 months would be reserved at the end for intensive nonresponse conversion by the traveling team of interviewers. The second phase was planned to overlap with the first, beginning at about month 3.

To allow more time to increase response rates, a decision was made late in the field period to extend the field period to approximately 40 weeks. Figure 8-1 shows the cumulative percentage of cases that were closed out, by week of the field period, for all household sample cases. The three-phase approach to data collection, as described in section 8.6.1, is shown here as well. (The spike in closed-out cases at the end of the field period occurred because nonresponse cases were not finalized until the end of the study.)

The pace of the field effort was influenced by several factors unique to NAAL. First, the design called for respondents to complete the background questionnaire, the assessment, and the oral module in the same visit; thus, it was necessary for a respondent to have a period of more than 1 hour during which he or she was reasonably unlikely to be interrupted. This requirement reduced the likelihood that respondents would be available on the interviewer's first visit and necessitated additional callbacks to ensure the completion of the case. In addition, because of the very large number of cases, across-the-board decisions to improve productivity took considerable time to implement.

Further, several natural disasters, such as hurricanes, fires, and snowstorms, made interviewing in selected areas impossible for periods of time. Finally, the climate of fear and suspicion fueled by the September 11, 2001 terrorist attacks created an environment in which people were less likely to participate.

**Figure 8-1. Percentage of closed-out cases, by week of field period and phases of data collection: 2003**



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

### 8.6.3 Data Collection Procedures

The interviewer’s first task was to conduct an in-person screening interview with a respondent at each of the sampled housing units that was occupied. Interviewers were instructed to administer the screener to a household member aged 16 or older, in order to identify eligible sample persons in the household.

Interviewers were required to complete the background questionnaire, the assessment (either the main assessment or the ALSA), and the oral module with the selected respondent in the same visit. The interviewer informed the respondent of the amount of time needed to complete all instruments and

attempted to transition directly into the interview after completing the screener. In situations where this was not possible, the interviewer set an appointment time to complete the interview and confirmed the appointment by telephone at a later date.

If a respondent demonstrated reluctance to participate, through either numerous broken appointments or a voiced refusal, the interviewer completed a noninterview report form and discussed further strategies with the supervisor.

A toll-free hotline was established for respondents to call with any additional questions (see section 3.3.6.3). The telephone number for this hotline was included on numerous study materials, such as the advance letter, brochure, letter of introduction, and community authorization letter. A total of 785 calls were received by the hotline over the course of the study. Most of the calls were requests for study verification, appointment scheduling, or refusals.

The average administration time for the NAAL instruments is displayed in table 8-3.

**Table 8-3. Average administration time, by instrument: 2003**

| Instrument                     | Average administration time<br>(minutes) |
|--------------------------------|--|
| Total                          | 91.8                                     |
| Main assessment sample persons |  |
| Screener                       | 4.5                                      |
| Background questionnaire       | 27.4                                     |
| Core and main assessment items | 45.7                                     |
| Oral module                    | 14.2                                     |
| Total                          | 77.1                                     |
| ALSA sample persons            |  |
| Screener                       | 4.5                                      |
| Background questionnaire       | 27.4                                     |
| Core and supplemental study    | 32.2                                     |
| Oral module                    | 13.0                                     |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

## **8.7 QUALITY CONTROL MEASURES AND FEEDBACK TO STAFF**

In addition to the CAPI range and logic checks and home office review of completed cases, the quality control measures described in the following sections were implemented to ensure high-quality

work in the data collection phase. The procedures were designed to assess the quality and completeness of data as collected and to provide timely feedback to the supervisors, the home office, and the interviewers.

### **8.7.1 Validation**

A 10 percent subsample of completed and noninterview cases were randomly selected for validation. Because falsification activity that goes undetected for a long time is very costly to correct, it is desirable to perform validation as soon after interviewing as possible. The validation efforts conducted by the field supervisors were supplemented by experienced interviewers from the data collection contractor's Telephone Research Center to ensure timely validation efforts.

Validation was performed by telephone when possible. On all cases selected for validation for which telephone numbers were not available, in-person validation was performed by a different interviewer who worked in the same or a nearby PSU.

For completed cases, the validation interview verified that contact had been made and confirmed the respondent's address at the time of the contact. Then three questions from the background questionnaire were asked again, to verify that the responses were consistent with those provided earlier. The respondent was also asked how long the interviewer spent with the respondent on the day of the interview, whether the interviewer assisted the respondent with the assessment, whether the interviewer conducted the oral module, and how much the respondent had been paid for participating in the study. For incomplete interviews, the validation interviewer verified contact (if the interviewer's original report of the case indicated that contact had occurred), confirmed the respondent's address and the number of household members at the time of contact, and, if possible, tried to schedule an appointment for an interview.

As soon as validation for a case was completed, a validation result code for the case was entered into the Data Management System. The Data Management System was used to monitor the progress of the validation effort and to ensure that at least 10 percent of each interviewer's work was being validated. At the end of data collection, 5,458 cases had been validated either by telephone or in person, for an overall validation rate of just over 15.4 percent of finalized screeners. Some of the validation workload in excess of the original 10 percent requirement resulted from situations in which additional verification was required to dismiss or confirm suspected falsification of interviews.

Of the nearly 400 field interviewers who worked on NAAL, 17 were discovered to have submitted some fraudulent work. Each of these interviewers was dismissed as soon as falsification was

confirmed, and all of their completed work was validated in person by another interviewer. Through in-person and telephone validation, it was determined that 263 of the 684 cases completed by these interviewers (38 percent) had been falsified. The data associated with these cases was deleted and the cases were reassigned to the field for completion.

### **8.7.2 Computer-Generated Reports**

The Supervisor Management System was used to manage and monitor the progress of the field work and provided critical management information to field and home office staff. One of the Supervisor Management System components was a reports mechanism. Reports were updated in real time as an interviewer or supervisor entered information or transmitted data.

The reports allowed all levels of management to monitor, on a daily basis, the progression of completion rates, response rates, and distribution of cases in pending codes by region, interviewer, PSU, and segment for each of the study instruments. Managers were also able to view the daily EROC information entered by interviewers. Costs were entered and monitored on a weekly basis as well.

The reports available to the management staff are discussed below.

The Data Collection Reports and the Interviewer Reports allowed the supervisors to view the overall status of production and response rates for their region and the nation.

The Interviewer Conference Report gave the supervisor a detailed view of pending cases for every interviewer, including the status for each component of the case. The Interviewer Cost Summary Report listed each interviewer's weekly and cumulative hours, expenses, and number of completed background questionnaires. Using these two reports as a starting point, supervisors discussed each pending case as well as overall completion rates, response rates, and costs with each interviewer during the weekly conference call. Plans and strategies for handling each pending case and overall performance were discussed during these calls, and workloads were adjusted as necessary. The quality of an interviewer's work was also discussed, based on the results of tape edits, interviewer observations, and validations.

The Missing Booklet Report listed finalized cases for which booklets had not been receipted in the field room.

The Interviewer Transmission Report documented each interviewer's last transmission of new data. Supervisors used this report to determine whether an interviewer was having difficulty in transmitting because of a problem in understanding the transmission process or a problem with the laptop. Supervisors monitored this report very closely because it affected the accuracy of all other reports.

The Assignment History of Nonfinalized Cases Report allowed supervisors to view and track the history and progress of reassigned cases. The supervisors could view, in one place, each interviewer's attempts to complete the instruments for cases that had been reassigned for efficiency and nonresponse conversion.

The Unassigned Cases Report documented any unassigned cases and also allowed the supervisor to check for entry errors in assigning or transferring cases.

The Validation Report documented all cases preselected for validation, including the completion or finalization date of each case to indicate that it was available for validation. Supervisors used this report to document the results of their validation efforts and to select any additional or substitute cases for validation. A summary report by regional and national totals was also available.

Additional reports were run each week to allow the project statisticians to monitor the sample yield for various populations in the national and state samples. Selected variables from the sample selection file, which carried census race characteristics by segment and by PSU, were merged with production data from the Supervisor Management System to allow a comparison of projected and actual results.

Reports including interview timing and scheduling data were also produced for each interviewer, from the CAPI database of finalized cases. Field managers used the reports to look for anomalies within the data that might identify interviewer falsification of interviews. Examples of these anomalies are very short instrument administration times, a short amount of time between interviews conducted at two different households, and interviews conducted very early in the morning or late in the evening.

### **8.7.3 Observation**

Two methods were used to observe NAAL interviews: (1) tape recording of interviews for review by supervisors and (2) in-person observations by home office staff and field managers and supervisors.

### **8.7.3.1 Tape-Recorded Interviews**

Supervisors relied on the review of tape-recorded interviews to “observe” each interviewer. Each interviewer was required to tape record two complete interviews, that is, the entire screener, background questionnaire, assessment, and oral module (to the extent that the respondent was able to complete the instruments). So that cases worked early in the data collection period could be reviewed and any errors in administration corrected, interviewers were asked to tape their 3rd and 20th interviews; if the respondent did not give consent, the next interview was taped instead. After listening to the tape, the supervisor completed a tape edit form and gave the interviewer feedback on the quality of the interviewing techniques and any mistakes or areas for improvement. A total of 569 tape-recorded interviews were evaluated.

### **8.7.3.2 In-Person Observations**

Interviewer observations were performed by the home office field managers; other members of the contractors’ and NCES staffs; and supervisors whose field offices were in sampled PSUs. Interviewer observations were performed for two main purposes. One purpose was to give home office staff an opportunity to observe respondents’ reactions to the study and also to observe how well the field procedures worked. Supervisors identified their strongest interviewers for this type of observation. The second purpose was to observe interviewers whose performance was of some concern, either because of their evaluation during training or because they were assigned to a particularly difficult area. Interviewers new to interviewing were observed as well.

Interviewers were typically observed locating sampled housing units, making screener contacts, setting appointments, and completing at least one background questionnaire, assessment, and oral module. During an interview, the observer listened but did not participate in any way. After the interview, when the observer and interviewer had left the respondent’s home, the observer used an interviewer observation form to evaluate the quality of the interviewer’s work. Interviewers were evaluated on the following points: ability to gain access to the household, organization of material and equipment, knowledge of the study, administration of the instruments, and general interviewing techniques. A total of 31 field observations were completed.

### **8.7.4 CAPI Help Desk**

A CAPI help desk was established and operated by staff specially trained in the NAAL instrumentation. If interviewers or supervisors experienced technical problems with the CAPI system,

they could call the toll-free help desk number and receive assistance in resolving the problem. The help desk received 1,575 calls during the data collection period.

The largest number of calls to the help desk related to data cleanup (cases where data required editing or cleaning by the home office); transmission, time and expense reporting, and shipment (e.g., connecting the laptop to the home office through the telephone lines, sending and receiving cases, and accessing or entering shipment data); laptop hardware and equipment (e.g., mouse, power cord, and carrying case); and the Interviewer Management System or e-mail (problems with user names or passwords, launching CAPI instrumentation, or receiving e-mail messages).

## **8.8 ACHIEVING HIGH RESPONSE RATES**

Response rates on household studies are influenced by three broad categories of factors:

- the ability of the interviewers to obtain cooperation;
- the effectiveness of “callback” procedures; and
- the efforts made by interviewers and supervisors to convert initial nonresponse cases to completed interviews.

These factors are described in the following subsections.

### **8.8.1 Interviewers’ Ability to Obtain Cooperation**

An important factor in maximizing response rates is the ability of the interviewers to encourage respondents to participate. Two sessions during the interviewer training program involved round table discussions and exercises focused specifically on techniques for handling reluctant respondents, answering questions, and avoiding refusals. Before working with actual households, the interviewers progressed through several stages of practice during training. During the first stage, they conducted role-playing exercises with one another until they felt comfortable and could demonstrate an adequate level of skill in gaining cooperation. After reaching this point, the interviewers conducted a practice interview with a paid volunteer respondent. Training staff monitored the interviewers closely during this segment of the training.

To assist the interviewers in gaining respondent cooperation, all sampled households received a cover letter and brochure approximately 7 to 10 days before the interviewer attempted to complete the screener.

## **8.8.2 Callback Procedures**

Developing an effective strategy for visiting housing units is a fundamental of good interviewing. Interviewers were trained in the following rules to build response rates:

- make trips at different times of day (morning, afternoon, or evening), taking into account that late afternoons and evenings would be the most productive hours in most cases;
- make trips on different days of the week; and
- make trips on weekends (Saturday or Sunday).

If they were unable to complete a screener during the first four visits to a housing unit, the interviewers were instructed to complete a noninterview report form and call the supervisor, who would either authorize more visits or assign the case to another interviewer.

## **8.8.3 Efforts to Convert Nonresponse**

Each type of nonresponse required a different strategy for conversion. The conversion strategies are summarized in the following sections.

### **8.8.3.1 Refusals**

Refusals are the most difficult type of nonresponse case to convert. When a respondent refused or broke off an interview, the interviewer completed the noninterview report form to capture information about the reason for the refusal. Using this information, the interviewer and supervisor could decide to send the respondent one of several refusal conversion letters or transfer the case to a different interviewer. Seven refusal conversion letters were developed for the main data collection:

- general refusal letter;
- no selling letter;
- too busy letter (in English and Spanish);
- sample person selected letter (in English and Spanish);
- elder letter (written in large print);
- locked building letter; and
- a simple, short letter designed for those who were not expected to take the time to read a more detailed letter.

### **8.8.3.2 Not at Home**

Interviewers were supplied with sorry-I-missed-you cards that could be left at the door when nobody was home. Interviewers were instructed to return to the home at different times of the day and on different days of the week to attempt to contact household members.

### **8.8.3.3 Language Problem**

Whenever possible, interviewers fluent in Spanish were sent to Spanish-speaking households. When a bilingual interviewer was not available, interviewers were instructed to locate a bilingual household member or neighbor or to hire a translator from a local community center to assist in conducting the screener component of the interview.

In households where neither English nor Spanish was spoken, interviewers tried to arrange for a household member, child, or neighbor (aged 16 or older) to assist in the translation, for the screener only. The background questionnaire, assessment, and oral module were not administered to respondents who could not speak English or Spanish.

### **8.8.3.4 Illness**

Whenever a respondent was too ill to participate, interviewers filled out a noninterview report form and discussed the situation with their supervisor.

### **8.8.3.5 Vacant Housing Unit**

If a housing unit was vacant during the interviewer's first visit, the case was closed out as vacant. If, however, the interviewer made initial contact with a household but returned to find that the housing unit was vacant or that a new family had moved in, the interviewer attempted to interview the household members who had lived in the housing unit at the time of the original contact.

## **8.8.4 Strategies To Increase Response Rates**

Many strategies were employed throughout the NAAL field period to increase response rates. They included collecting information about the sampled households from neighbors, administering a hard-copy version of the screener, sending FedEx mailings, and implementing interviewer incentive programs. These strategies are discussed below.

#### **8.8.4.1 Neighbor Information**

After unsuccessful attempts to contact the sampled address, interviewers were allowed to use neighbor information as a mechanism to identify ineligible housing units in high-minority segments only. Interviewers were required to collect the following information from two neighbors of the selected household: whether the housing unit was occupied, the best time and day to contact the household, and whether any household members were Hispanic or Black. The names, addresses, and telephone numbers of the neighbors were collected as well. If neither neighbor report indicated that there were minority members in the sampled household, the supervisor finalized the case as ineligible. In all other circumstances, the interviewer was required to continue efforts to interview the sampled household. This procedure was implemented in November 2003 and used for the remainder of the field period.

#### **8.8.4.2 Hard-Copy Screener**

To increase response rates, a hard-copy version of the CAPI screener was developed and implemented in segments where a hard-copy screener was deemed preferable to administering a CAPI instrument at the doorstep. This approach was used in households where the presence of the laptop computer was expected to be intimidating or imposing to potential respondents. Once the data had been collected on the hard-copy form, NAAL interviewers transferred the information into the CAPI screener to determine whether any household members were eligible for the study.

The hard-copy screener was used by the NAAL interviewers, as well as by approximately 55 experienced interviewers on short-term loan from other studies conducted by the data collection contractor. The non-NAAL interviewers were teamed with a NAAL interviewer who followed up with the selected respondent to complete the remainder of the interview, if an eligible respondent was identified. Nearly 1,000 screeners were completed by these non-NAAL interviewers.

#### **8.8.4.3 FedEx Mailings**

In December 2003 and January 2004, more than 6,000 FedEx mailings were sent to households where interviewers had made numerous visits but were unable to find anyone at home and where attempts at refusal conversion had been unsuccessful. The FedEx mailings contained the NAAL advance letter, the study brochure, and a brief letter, when appropriate. These materials helped introduce the interviewer to households that had not received or read the previous mailings.

#### **8.8.4.4 Interviewer Incentives**

Several interviewer incentive plans aimed at increasing production were put into place throughout the last several months of the data collection period. These plans encouraged interviewers to complete as many interviews as possible during periods when they were less likely to work, such as harsh weather conditions and the winter holiday season. These additional incentive plans were found to spur production. All interviews completed as part of an incentive plan were validated.

#### **8.8.2 Reasons for Nonresponse**

There are two major reasons for nonresponse: (1) literacy-related nonresponse,<sup>2</sup> that is, sample persons who did not respond because of language problems, reading or writing barriers, or mental disabilities; and (2) other nonresponse, that is, sample persons who did not respond for other reasons (e.g., refusals or unlocatable cases). Table 8-4 shows the sample counts and percentages by reason for nonresponse for the national and state samples for the screener, the background questionnaire, the assessment, and the oral module. Among these instruments, literacy-related nonresponse was most prevalent for the background questionnaire (1.5 percent of sample persons).

Of the 35,365 housing units sampled in the national and state components, 13.2 percent were unoccupied; that is, they were either vacant (8.6 percent) or were not a housing unit (4.6 percent). A screener was completed at 25,123 of the 30,694 occupied housing units in the sample, for an overall unweighted screener response rate<sup>3</sup> of 81.8 percent. The largest category of screener nonresponse was “refusals,” which occurred in 3,207, or 9.1 percent, of the occupied households. The second largest category was maximum callbacks, which occurred in 1,625, or 4.6 percent, of the occupied households. There were 160 cases of literacy-related nonresponse (0.5 percent), all classified as language problems.

The screening effort identified 22,270 households that included at least one eligible person. Among the 23,732 sample persons, 18,186 (76.6 percent) completed the background questionnaire. Refusals, which accounted for the largest number of nonrespondents to the background questionnaire, occurred in 3,032 cases (including 153 cases where someone refused for the sample person), or 12.7 percent of all sample persons. The next category was maximum callbacks, which accounted for 1,401

---

<sup>2</sup> Of the 18,541 cases that were weighted in the household sample, 439 did not complete the assessment due to mental disability, language problem or reading/writing barrier. These literacy-related cases are counted as respondents for response rate and nonresponse bias analysis purposes, since some useful data was collected from the screener and background questionnaire (for some), and some knowledge of the English literacy skills was also obtained. After the weighting of the data and the computation of response rates, a decision was made to exclude disabilities and cases with reading/writing barriers from the target population. However, these cases are in scope for this report.

<sup>3</sup> Weighted response rates are discussed in chapter 11.

cases (5.9 percent). A total of 355 sample persons (1.5 percent) were considered literacy-related nonrespondents because of a language problem (211 cases, or 0.9 percent) or mental disability (144 cases, or 0.6 percent).

Table 8-4 also shows the reasons for nonresponse to the assessment. Of the 23,732 sample persons, 17,172 (72.4 percent) completed the assessment, and an additional 548 (2.3 percent) partially completed it. The reasons for partial completion were both literacy related (1.0 percent) and nonliteracy related (1.3 percent). Further, 5,540 sample persons did not attempt the assessment because they did not complete the background questionnaire. Of the 6,012 sample persons who did not attempt the assessment, 290, or 1.2 percent of all sample persons, were classified as refusals.

Only 51 sample persons (0.2 percent) were classified as having literacy-related reasons for not attempting the assessment, including language problems (27 cases, or 0.1 percent), reading/writing barriers (14 cases, or .1 percent), and mental disabilities (10 cases, or 0 percent).

The reasons for nonresponse to the oral module are also provided in table 8-4. Of the 23,732 sample persons, 17,057 (71.9 percent) completed the oral module, and an additional 23 (0.1 percent) partially completed it. The reasons for partial completion were both literacy related (9 cases) and nonliteracy related (14 cases). Further, 6,199 sample persons did not attempt the oral module because they did not complete the assessment. Of 6,652 sample persons who did not attempt the oral module, 191, or 0.8 percent of all sample persons, were classified as refusals. Only 145 sample persons (0.6 percent) were classified as having literacy-related reasons for not attempting the oral module, including language problems (70 cases, or 0.3 percent), reading/writing barriers (50 cases, or 0.2 percent), and mental disabilities (25 cases, or 0.1 percent). For response rates associated with the household, by major instruments and by key variable, refer to chapter 11.

**Table 8-4. Sample counts by instrument and reasons for nonresponse: 2003**

| Instrument and reasons for nonresponse              | National household total | State household total | Combined household total |
|---|--------------------------|-----------------------|--------------------------|
| <b> Screener</b>                                    |                          |                       |                          |
| Total housing units                                 | 25,450                   | 9,915                 | 35,365                   |
|   | Percent                  | Percent               | Percent                  |
| All housing units (percent)                         | 100.0                    | 100.0                 | 100.0                    |
| Complete, 1 sample person                           | 56.2                     | 65.6                  | 58.8                     |
| Complete, 2 sample persons                          | 4.1                      | 4.1                   | 4.1                      |
| Complete, subsampled out using neighbor information | 0.6                      | 0                     | 0.4                      |
| Complete, subsampled out                            | 10.6                     | 0                     | 7.6                      |
| Illness   | 0.5                      | 0.8                   | 0.6                      |
| Language problem                                    | 0.4                      | 0.5                   | 0.5                      |
| Maximum callbacks                                   | 4.9                      | 3.8                   | 4.6                      |
| Not a housing unit                                  | 4.6                      | 4.6                   | 4.6                      |
| Other   | 0.9                      | 0.8                   | 0.8                      |
| Refused   | 8.2                      | 11.2                  | 9.1                      |
| Unavailable for field period                        | 0.2                      | 0.2                   | 0.2                      |
| Vacant  | 8.8                      | 8.3                   | 8.6                      |
| <b> Background questionnaire</b>                    |                          |                       |                          |
| Total sample persons                                | 16,409                   | 7,323                 | 23,732                   |
|   | Percent                  | Percent               | Percent                  |
| All sample persons (percent)                        | 100.0                    | 100.0                 | 100.0                    |
| Complete  | 76.1                     | 77.8                  | 76.6                     |
| Language problem                                    | 1.0                      | 0.6                   | 0.9                      |
| Maximum callbacks                                   | 6.5                      | 4.5                   | 5.9                      |
| Mental disability                                   | 0.6                      | 0.6                   | 0.6                      |
| Other   | 0.8                      | 0.7                   | 0.8                      |
| Physical disability                                 | 1.4                      | 1.3                   | 1.4                      |
| Sample person refused                               | 11.8                     | 12.8                  | 12.1                     |
| Someone refused for sample person                   | 0.7                      | 0.6                   | 0.6                      |
| Unavailable for field period                        | 1.1                      | 1.0                   | 1.0                      |

See notes at end of table.

**Table 8-4. Sample counts by instrument and reasons for nonresponse: 2003—Continued**

| Instrument and reasons for nonresponse                    | National household total | State household total | Combined household total |
|---|--------------------------|-----------------------|--------------------------|
| <b>Assessment</b>   |                          |                       |                          |
| Total sample persons                                      | 16,409                   | 7,323                 | 23,323                   |
|   | Percent                  | Percent               | Percent                  |
| All sample persons (percent)                              | 100.0                    | 100.0                 | 100.0                    |
| Reading/writing barrier                                   | 0.1                      | 0.1                   | 0.1                      |
| Complete  | 71.6                     | 74.2                  | 72.4                     |
| Language problem  | 0.1                      | 0.1                   | 0.1                      |
| Mental disability   | 0.1                      | 0                     | 0                        |
| Not applicable, did not complete background questionnaire | 23.8                     | 22.2                  | 23.3                     |
| Other   | 0.1                      | 0.1                   | 0.1                      |
| Partial complete, reading/writing barrier                 | 0.6                      | 0.2                   | 0.5                      |
| Physical disability                                       | 0.4                      | 0.3                   | 0.4                      |
| Partial complete, refused, core incomplete                | 0.2                      | 0.1                   | 0.2                      |
| Partial complete, language problem                        | 0.5                      | 0.2                   | 0.4                      |
| Partial complete, mental disability                       | 0.1                      | 0.1                   | 0.1                      |
| Partial complete, physical disability                     | 0.3                      | 0.8                   | 0.5                      |
| Partial complete, refused, core complete                  | 0.7                      | 0.4                   | 0.6                      |
| Sample person refused                                     | 1.2                      | 1.0                   | 1.2                      |
| Someone refused for sample person                         | 0                        | 0.1                   | 0.0                      |
| Unavailable for field period                              | 0                        | 0.1                   | 0.1                      |
| <b>Oral module</b>  |                          |                       |                          |
| Total sample persons                                      | 16,409                   | 7,323                 | 23,732                   |
|   | Percent                  | Percent               | Percent                  |
| All sample persons (percent)                              | 100.0                    | 100.0                 | 100.0                    |
| Reading/writing barrier                                   | 0.2                      | 0.2                   | 0.2                      |
| Complete  | 71.0                     | 73.8                  | 71.9                     |
| Language problem  | 0.4                      | 0.1                   | 0.3                      |
| Mental disability   | 0.1                      | 0.1                   | 0.1                      |
| Not applicable, did not complete assessment               | 26.8                     | 24.5                  | 26.1                     |
| Other   | 0.1                      | 0.2                   | 0.1                      |
| Partial complete, reading/writing barrier                 | 0                        | 0                     | 0                        |
| Partial complete  | 0                        | 0                     | 0                        |
| Physical disability                                       | 0.3                      | 0.5                   | 0.4                      |
| Partial complete, language problem                        | 0                        | 0                     | 0                        |
| Partial complete, mental disability                       | 0                        | 0                     | 0                        |
| Partial complete, physical disability                     | 0                        | 0                     | 0                        |
| Sample person refused                                     | 0.9                      | 0.6                   | 0.8                      |
| Unavailable for field period                              | 0                        | 0                     | 0                        |

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

### 8.8.3 Demographic Profiles of Respondents

This section presents demographic profiles of the respondents at each stage. Table 8-5 presents the distributions of respondents with respect to selected demographic and other characteristics, for the screener, background questionnaire, assessment, and oral module. The counts and percentages presented in the table are unweighted and were computed from the combined NAAL-SAAL household sample.

Table 8-5 shows some simple distributions among the samples of respondents at the various stages of data collection. For a comprehensive analysis of the differences between respondents and nonrespondents, a complete nonresponse bias analysis is provided in chapter 11. The nonresponse bias analysis provides an evaluation of the potential for bias owing to nonresponse to the screener and the background questionnaire.

Table 8-5 also provides the sample counts of screener respondents. The respondent counts reflect the number of housing units that completed the screener or were ineligible because of subsampling procedures. Minority status was defined as high if the sample housing unit was from a segment in which at least 25 percent of the population was Black or Hispanic and was defined as low otherwise. The majority of respondent housing units came from the South and were from metropolitan statistical areas (MSAs).

The background questionnaire respondent counts reflected both the number of sample persons who completed the background questionnaire and the number who did not complete the background questionnaire for literacy-related reasons. About 36 percent of the sample were minority sample persons (i.e., Hispanic or non-Hispanic Black adults), which was a reflection of the oversampling procedures. The distributions of respondents for the assessment and the oral module were very similar to the distribution for background questionnaire respondents because the level of nonresponse to the assessment and oral module was low among background questionnaire respondents. It should be noted that assessment and oral module respondents were defined as those who (1) completed the assessment; (2) those who only partially completed the assessment (because of a language problem, mental disability, physical disability, or reading/writing barrier); and (3) those who did not attempt the assessment (or oral module) because of a language problem, mental disability, physical disability, or reading/writing barrier.

**Table 8-5. Demographic profile of respondents in the combined household sample, by instrument: 2003**

| Characteristic                     | Screener total | Background questionnaire total | Assessment total | Oral module total |
|------------------------------------|----------------|--------------------------------|------------------|-------------------|
|                                    | Percent        | Percent                        | Percent          | Percent           |
| Total sample                       | 25,123         | 18,541                         | 17,668           | 17,300            |
| Total percent                      | 100.0          | 100.0                          | 100.0            | 100.0             |
| <b>Region</b>                      |                |                                |                  |                   |
| Northeast                          | 19.6           | 20.3                           | 20.2             | 20.2              |
| Midwest                            | 18.4           | 19.5                           | 19.4             | 19.4              |
| South                              | 44.3           | 44.6                           | 45.3             | 45.6              |
| West                               | 17.7           | 15.6                           | 15.0             | 14.9              |
| <b>MSA<sup>1</sup> status</b>      |                |                                |                  |                   |
| MSA                                | 80.6           | 79.0                           | 78.7             | 78.8              |
| Non-MSA                            | 19.4           | 21.0                           | 21.3             | 21.2              |
| <b>Minority status<sup>2</sup></b> |                |                                |                  |                   |
| High                               | 45.7           | 40.4                           | 40.3             | 40.2              |
| Low                                | 54.3           | 59.6                           | 59.7             | 59.8              |
| <b>Gender</b>                      |                |                                |                  |                   |
| Male                               | †              | 43.3                           | 42.9             | 42.8              |
| Female                             | †              | 56.7                           | 57.1             | 57.2              |
| <b>Age</b>                         |                |                                |                  |                   |
| 16–29                              | †              | 25.4                           | 25.8             | 26.0              |
| 30–49                              | †              | 39.2                           | 39.3             | 39.4              |
| 50–69                              | †              | 24.6                           | 24.6             | 24.6              |
| 70+                                | †              | 10.8                           | 10.3             | 10.1              |
| <b>Race/ethnicity</b>              |                |                                |                  |                   |
| Hispanic                           | †              | 17.2                           | 17.1             | 17.0              |
| Black                              | †              | 18.9                           | 19.2             | 19.2              |
| Other                              | †              | 63.9                           | 63.7             | 63.8              |

† Not applicable.

<sup>1</sup> Metropolitan Statistical Area.

<sup>2</sup> If a sampled segment (group of census blocks) contained 25 percent or more Black and Hispanics, then the segment was classified as high minority. Otherwise the segment was classified as low minority.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

## **8.9 DATA PREPARATION AND PROCESSING**

During the data collection period, interviewers returned materials to the home office in two formats: hard copy and electronic. The hard-copy materials returned included assessment booklets, ALSA questionnaires, household folders, noninterview report forms, and oral module interviewer guides. The electronic data consisted of screener, background questionnaire, assessment interviewer guide, oral module, and interviewer observation data, as well as status code updates, and data indicating which hard-copy materials had been shipped by the interviewer to the home office.

The Data Management System was used to support data processing activities, among other functions. The Data Management System consisted of three main components: the Supervisor Management System (see section 8.3.2.2), receipt control functions, and the reporting mechanism. This section of the report focuses on the receipt control functions.

During the field period, interviewers typically transmitted interview data to the home office daily. During transmission, the updated status codes and questionnaire data for all completed and in-process cases were combined and sent to a server through a secure dial-up connection. Transmitted files were backed up and held on the server. Approximately every 5 minutes, an automatic process determined whether new transmissions had been received from the field interviewers. This process was an ongoing operation during the field period.

The process that handled data transmissions performed two functions: It updated the Data Management System database with case status information, and it moved the completed interview data to processing directories on the project server. Once a day, the newly transmitted Blaise data files created during the interview were decrypted and concatenated to create a project-level Blaise database for each instrument.

The study database was updated almost immediately after the transmitted data were received. Backup processes were in place to ensure that transmitted data were received successfully. During the conversion of the Blaise data to the study database, two other products were created. The first was a Blaise interview browse area that allowed project staff to locate data on individual cases. This interview browse function was used primarily in resolving issues reported to the hotline. The second product was a SAS data set that was used to complete ad hoc reporting requests, such as task timing reports.

### **8.9.1 Overview of Data Preparation and Processing Activities**

Several data preparation and processing systems and procedures were developed to support the extraction of screener, background questionnaire, and interviewer guide data from the Blaise system and into SAS data files. Cases transmitted by interviewers were appended to a cumulative database, where they were then prepared for a review of interviewer comments and coding of other-specify and open-ended responses. Once these steps were completed, the data were converted into the study database. Frequencies were run daily on the study database and reviewed for outliers and inconsistencies. After the frequency review task, the data were loaded into the Blaise Editing System, batched, tracked, and cleaned.

### **8.9.2 Receipt, Batching, and Editing of Cases**

Once data had been moved into the Blaise Editing System, manageable batches of 100 cases were created. The Blaise Editing System tracking file, which contained all the records within a batch, was then developed. A sequential batch number was used to link the tracking file records to the batch file records. After batching, a utility in the Blaise Editing System assigned batches to a data editor.

The Blaise Editing System management report identified the cases within each batch that required data cleaning. During the creation of the batches, any automated edits that had been triggered and suppressed by the interviewer during the interview were reinstated and the interview status was reset. Additional postcollection editing, including edits considered too time-consuming or complex to be executed in an interview situation, was conducted as well. The cases were then verified, coded, adjudicated, and prepared for delivery.

### **8.9.3 Processing of Hard-Copy Assessment Materials**

Interviewers were instructed to return completed interview materials (including assessment and ALSA booklets, household folders, oral module interviewer guides, and noninterview report forms) to the home office twice a week. All shipments were tracked electronically with the Interviewer Management System and the Data Management System. Before the interviewer mailed the items, the electronic shipping record section of the Interviewer Management System was used to indicate each item included in the shipment. After marking each case and item being sent, the interviewer entered the shipping tracking number into the Interviewer Management System. The home office was then able to see which case materials were in transit.

The status and location of all assessment and ALSA booklets were constantly monitored. The Data Management System automatically changed the status associated with each booklet to reflect when it was sent to an interviewer, when it was marked as shipped by the interviewer, and when it was received at the home office. Upon arrival, completed interview materials were receipted in the Data Management System, and items received were compared against the list of items expected. Any discrepancies were brought to the attention of the systems management team and, where needed, the field director and supervisors.

Once an assessment booklet had been receipted, it was filed by case ID and, where appropriate, sample person ID within case. ALSA booklets, household folders, and oral module interviewer guides were filed separately by case ID. Noninterview report forms were maintained in the associated household folder.

As part of the standard NAAL quality control procedures, all seven core items in the completed assessment booklets were rescored by trained home office staff. The validation results were entered into a specially designed core scoring program. The program compared the interviewer's scoring with that of the home office staff, enabling supervisors to give interviewers feedback on their performance.

Early in the data collection period, the core items were rescored for 100 percent of the receipted assessment booklets. When interviewers were determined to be proficient at scoring the core items, no further core validation was conducted for those interviewers. Home office staff continued to rescore 100 percent of the core items for the remaining interviewers.

The in-house validation of the core items continued through the middle of January 2004, as the end of data collection approached. Home office staff rescored a total of 13,608 core assessments. (See section 13.3.2.2 for a further discussion of core validation.)

#### **8.9.4 Processing of Oral Module Data and Zip Disks**

Interviewers were required to mail a zip disk of oral module recordings to the home office once each week. Each disk contained all the oral module interviews completed by the interviewer during the week. The files received from the interviewers were copied into a backup directory. These files were maintained until confirmation was received from the oral module scoring contractor that the recordings had been successfully received and processed. Zip disks were sent for scoring on a weekly basis throughout the data collection period.

### **8.9.5 Data Processing for the ALSA**

As discussed in section 8.9.3, ALSA booklets were receipted in the home office. Following receipt, the booklets were reviewed and edited, during which trained staff reviewed notes written in the margins and ensured that every questionnaire item had a valid response associated with it and that the skip patterns had been correctly followed by the interviewer. After the booklets had been edited, they were prepared for data entry. All booklets were double-keyed by independent data entry staff at a rate of 100 percent. Any discrepancies between the two entry processes were reconciled. After the data entry procedure was completed, the codebooks and associated frequencies were reviewed and adjudicated.

### **8.9.6 Delivery of Assessment Booklets to the Scoring Contractor**

Assessment booklets were delivered to the assessment scoring contractor at four points during the field period. Each shipment included an electronic file listing all the items included in the shipment for verification purposes.

### **8.9.7 Quality Control of Data in the Study Database**

As discussed above, a detailed reporting system, integrated with the Data Management System, was used throughout the main study data collection. Data in these reports were reviewed by the field director, field managers, supervisors, and home office staff throughout the field period.

In addition, receipt control reports were used to track and verify that assessment booklets were sent to the home office on a timely basis after completion. The integrated receipt control system also tracked the assessment booklets throughout the field period and documented the location of each booklet.

At the end of the field period, an extensive data validation process was completed to ensure that Blaise interview data existed for each completed case in the Data Management System. Status codes were compared to ensure that each status within the Data Management System accurately reflected the existence of the Blaise interview data. The CAPI instrument data were compared, and all discrepancies were documented, reviewed, and corrected, if feasible to do so, before delivery of the final data to the analysis contractor. In addition, some problem cases were restored from laptop backup disks to ensure the existence of data for all completed cases.

Additionally, the assessment scoring data were reviewed to ensure that a score had been received for each case ID and booklet barcode number sent for scoring. When discrepancies could not be resolved, differences were documented in the final delivery documents.

Finally, a thorough process of reconciling FAN data was conducted to ensure that data were available for all finalized cases with a completed FAN interview. Any cases that were determined to be missing from the scoring database were reprocessed. The associated responses were then generated and the data were redelivered.

*This page is intentionally left blank.*

## **CHAPTER 9**

### **CORRECTIONAL INSTITUTION STUDY DATA COLLECTION AND QUALITY CONTROL**

Michelle Amsbary, Terri Annis, and Martha Berlin, Westat

#### **9.1 INTRODUCTION**

The purpose of the Correctional Institution Study, hereinafter referred to as the Prison Study, was to assess the literacy skills of adult inmates in federal, state, and private correctional facilities, using a sample of approximately 1,100 inmates. The inclusion of the inmate sample in the National Assessment of Adult Literacy (NAAL) helped improve estimates of the literacy levels of the total U.S. population and made it possible to report on the proficiencies of this important segment of society. The Prison Study component was developed in consultation with the Bureau of Justice Statistics of the U.S. Department of Justice and the Federal Bureau of Prisons.

Data collection took place following the completion of the main household study. The field period ran for approximately 4 months, from mid-March through July 2004. A total of 1,173 background questionnaire interviews were completed in 107 facilities. A complete description of the sample design is provided in chapter 7.

To ensure comparability with the NAAL household study, the inmates completed the same literacy tasks as the household population. However, to address issues of particular relevance to the prison population, a revised version of the background questionnaire was developed. New questions included queries about current offenses, criminal history, and participation in prison programs, as well as education and workforce experience. (See section 2.3 for a complete description of questionnaire development and content.)

As in the Prison Study conducted as part of the 1992 National Adult Literacy Survey (NALS), the rules of almost all facilities precluded monetary or in-kind incentives. Instead, a personalized certificate of participation was placed in the inmate's file upon completion of the interview.

Before the main study data collection, a small pretest was conducted at three correctional institutions, one in Maryland and two in Texas. Additionally, the Spanish version of the background

questionnaire was tested at a Maryland facility. The pretest evaluated the ease of administration of the study instruments, administration time, within-facility procedures, and inmate reaction to the study. On the basis of the pretest experience, minor changes were made to the background questionnaire to facilitate administration, and administrative procedures were refined to reflect lessons learned.

## **9.2 GAINING COOPERATION**

The permission and cooperation of federal, state, and correctional facility officials were required before interviewing in prisons. Representatives from the Bureau of Justice Statistics and from the Office of Vocational and Adult Education within the U.S. Department of Education provided considerable assistance in gaining cooperation from federal and state correctional agency officials. Letters of endorsement were obtained from the Correctional Education Association and the American Correctional Association. All of these organizations and individuals contributed to the success of the negotiation process. Of the 110 eligible facilities originally selected for the study, 107 (97.3 percent) agreed to participate.

The following steps were used to gain cooperation at the sampled facilities:

- The data collection contractor mailed letters to the correctional agencies of all states in which prisons had been selected for the study. A letter was also mailed to officials at the Federal Bureau of Prisons. The letter explained the study and asked for permission to contact selected facilities within the agency's jurisdiction. Letters were followed up with telephone calls to answer questions, secure cooperation, and determine prison contact procedures. For the federal prisons and several state institutions, the study protocol and instruments had to be approved by individual institutional review boards.
- The state or federal official, in most cases, informed the warden at the sampled facility that the facility had been selected and urged the facility to participate. The data collection contractor then contacted the facility. The contractor's prison negotiator provided additional information about the study and described the sample selection process. The warden was asked to approve the study protocol and to designate a prison official to serve as coordinator for the study. The prison negotiator and the designated coordinator then worked out details, such as the interviewing procedures within the facility.
- The interviewers assigned to conduct interviews at a facility contacted the prison coordinator 2 days before the scheduled sampling date to reconfirm negotiated arrangements, including production of the inmate list for sampling, and to resolve any outstanding details.

Facility negotiations included (1) procedures for providing interviewer security within the institution and (2) interviewer clearance procedures required by the facility. Prison coordinators were

asked to arrange a secure, private room for each interview. If this arrangement was not possible, interviews were conducted in partitioned or private areas of larger rooms where the inmate would not be interrupted and could be assured of confidentiality. Depending on the security regulations of each facility, respondents were either brought to the interview session by a guard or received a pass to meet with the interviewer unescorted. To minimize misinformation and deter refusals, facilities were requested to “call out” selected inmates without providing an explanation of the study. The interviewer was responsible for introducing the study and gaining inmate cooperation.

The interviewer clearance process varied from state to state and facility to facility. In three states, department of corrections officials conducted interviewer background checks and notified the facilities that the interviewers had been cleared. Elsewhere, each facility had its own requirements. In general, the interviewer’s name, address, Social Security number, date of birth, and driver’s license number were submitted to the facility and were processed by a recognized clearance agency. Study materials were generally reviewed by prison officials during the negotiation process. In some cases, special permission was required to allow interviewers to bring in the laptop used to administer the interview. Most facilities also required that interviewers obtain a signed informed consent form from inmates before the interview. The form included statements on confidentiality and assurances that participation or nonparticipation would not affect release or parole eligibility.

### 9.3 DATA COLLECTION MATERIALS AND INSTRUMENTS

The materials used in the conduct of the Prison Study are listed below:

- **Facility folder:** This was similar to the segment folder used in the household study (see section 8.2.2). A unique folder was developed for each facility and contained the facility scheduling sheet, facility sampling form, and sample listing sheet (described below).
- **Facility scheduling sheet:** This document contained information concerning the facility, including its location, contact information, logistical information, and security arrangements. Interviewers were required to review this form very thoroughly prior to visiting the facility because it included information that was unique to each facility. In addition to contact information, the facility scheduling sheet included details on where, when, and how inmate sampling would be conducted.
- **Facility sampling form:** This form was used in conjunction with a module in CAPI to select the inmates for participation (see section 9.4).
- **Sample listing sheet:** This sheet was used to schedule interviews and document the results of the interviews for each selected inmate. This form was provided to the facility

contact to aid in scheduling the interviews and ensuring that the appropriate inmates were available at the indicated time.

- **Screener:** The screener instrument was much abbreviated from that used in the household component of the study. The Prison Study screener collected simply the sample person's first name and institutional identification (ID) number, as assigned by the facility.
- **Background questionnaire:** While essentially the same instrument as administered in the household component, some of the questions were changed, deleted, or added to adapt the interview and information gathered to the correctional environment.
- The  **assessment, supplemental study, and oral module** were exactly the same as in the household component.
- **Certificate of participation:** Following the completion of the entire interview, inmates received a certificate of participation for their files. (In some cases, institutions did not permit the certificates; in other institutions, the certifications had to be put directly into the inmates' files.)
- **Prison noninterview report form (NIRF):** If a sampled inmate did not complete any component of the interview, the interviewer documented the reasons on the prison noninterview report form.
- **Participant folder:** Each sample person was assigned a participant folder. All contact results were recorded in the Record of Actions on the participant folder.

The materials discussed above were developed specifically for the Prison Study and the use of the materials are discussed in the subsequent sections.

## 9.4 INTERVIEWER SELECTION AND TRAINING

To conduct the data collection, 43 interviewers were recruited from among the NAAL household study workforce. Criteria for selection included proximity to sampled facilities, experience in interviewing in correctional facilities, availability, and willingness to interview in correctional facilities. The interviewer manual documented procedures specific to interviewing the prison population, including instructions for the facility contact and sampling forms, question-by-question specifications for each instrument, and instructions for reporting information. Training materials focused on the following aspects of the study:

- the background and purpose of the study, including an overview of facility negotiations (see section 9.2 for more details on gaining cooperation of the prisons);
- inmate sampling forms and procedures;

- specifications for administering the background questionnaire;
- procedures for working within correctional facilities, obtaining inmate cooperation, and reporting results of the study; and
- administrative procedures.

Interviewer training was conducted during a 2-day in-person session. Particular emphasis was given to the inmate sampling procedures, including numerous hands-on practice exercises. Interviewers also received information on security procedures and working in the correctional environment. Further details are provided in the following sections.

### **9.4.1 Sampling Procedures**

A significant portion of the training program focused on the sampling procedures for randomly selecting inmates in each of the correctional facilities. Much of the training time was devoted to the procedures involved in the creation of a sampling frame, or a list of all inmates who were eligible for the study, and selection of inmates. Although the facilities provided the inmate lists (based on detailed information provided by the study recruiters), interviewers were required to verify that the lists were current and contained all eligible inmates, and in some cases, arrange the lists into a suitable sampling frame.

The inmate lists provided by the facilities differed significantly in structure, format and content, based on the facility's size and record-keeping system. Interviewers were trained on how to accommodate and work with the various list structures through numerous training exercises.

Interviewers were required to follow the steps outlined on the facility sampling form to verify that the list of inmates provided by the facility included the following:

- all inmates who had a bed assigned and slept at the facility the previous night;
- all inmates who had a bed assigned but were temporarily absent (on furlough, in court, in the hospital, or in another facility); and
- all inmates who were admitted the previous day and had a bed assigned as of the previous night.

After adding the name and/or prison ID number of any eligible inmates who did not previously appear on the list, interviewers ensured that the list did not include any of the following:

- any inmates who were released from the facility prior to the previous night;
- any inmates who ordinarily slept elsewhere (hospital, halfway house, work release center, local jail);
- any inmates who were admitted to the facility after the previous night;
- any inmates who were scheduled to be admitted but were not officially at the facility the previous night;
- any inmates who had escaped or were away without leave;
- any inmates who were sleeping, eating, working, visiting, or for any other reason in the facility the previous night but did not ordinarily have a bed assigned there;
- any inmates who were released from the facility prior to the previous night; or
- any inmates who were under the age of 16.

If any of these inmates were found to be on the list, interviewers were trained to delete their names from the list. Once the list was determined to contain all eligible inmates, it was necessary for the interviewers to conduct a final count of all inmates on the list. A few strategies were developed to assist the interviewers in the task of manually counting the inmates on the list, and these strategies were practiced during the training session. Once the final count of inmates was written on the facility sampling form, a CAPI module was utilized to select a sample of inmates for interviewing.

After the interviewer input several pieces of information regarding the facility, including the final count of inmates, the CAPI module indicated how many were sampled and their corresponding line numbers from the lists. The final sampling task for interviewers was to determine which inmate name/ID corresponded with the line number selected by the CAPI system—these were the inmates who would be interviewed. Interviewers practiced these procedures through five in-depth, realistic training exercises. After selecting the sample at the first two facilities, each interviewer was required to contact the home office sampling coordinator by telephone to review the sampling process and ensure that the proper procedures were followed before scheduling interviews with the sampled inmates. The sampling coordinator remained available to discuss questionable or problematic sampling procedures for the remainder of the field period.

## **9.4.2 Inmate Participation**

The interviewer training session also covered procedures for scheduling the interviews and obtaining respondent cooperation within a correctional institution. Once the sampling procedure was completed and the sample listing sheet filled out, the sheet was provided to the facility contact who then determined the actual interview date, time, and location for each inmate and added the relevant information to the sheet. Interviews were scheduled a minimum of 2 hours apart.

Arrangements for an appropriate interview setting were made with the facility during the initial negotiation process (see section 9.2). Most interviews took place in administrative offices, attorney/client rooms, or classrooms. Facility contacts were asked not to discuss the NAAL study with the inmates prior to bringing them to the scheduled interview. Interviewers presented inmates with a letter of introduction upon arriving at the interview. The interviewer also discussed the certificate of participation and the informed consent form, as appropriate, before beginning the interview.

## **9.4.3 Background Questionnaire**

The general format and structure of the Prison Study background questionnaire was identical to that used in the household component of NAAL. Many of the questions were also identical, but some were deleted and others, which pertained directly to the prison environment, were added. The entire background questionnaire instrument was covered in training, with particular emphasis given to the prison-specific items, such as questions about current offenses, criminal history, and participation in prison programs, as well as education and workforce experience.

## **9.4.4 Administrative Procedures**

The final part of training focused on administrative procedures such as dismissing inmates at the end of the interview, presenting the certificate of participation, accounting for all interview materials, special security situations, recordkeeping, the assignment of result codes, and completing the prison noninterview report form. The administrative procedures and recordkeeping forms remained essentially the same as those employed in the household component, but interviewers were trained in the techniques needed to apply them to the Prison Study.

Interviewers were trained in the procedures specific to working within correctional facilities. For instance, correctional institutions had strict dress codes that interviewers had to be aware of prior to their visit. Additionally, many facilities had items they considered contraband, including food containers, tobacco products, chewing gum, and paper clips.

Security procedures were another focus of the training. Interviewers had to be prepared to undergo numerous security procedures at the facility, in addition to the background checks conducted prior to their being hired for the study. Facilities required numerous security procedures, ranging from requests for identification and use of facility visitor ID badges to physical searches such as metal detectors.

## **9.5 DATA COLLECTION**

On average, 4 days were required to select the sample of inmates and administer the interviews in each facility. Two interviewers usually entered the facility on Monday morning to sample inmates and submit the list of selected inmates to the prison coordinator for scheduling interview appointments. Interviewing typically began on Tuesday and proceeded at the rate of about four interviews a day. Interviewer assignments were guided by the proximity of the interviewer's home to the facility, by the need for bilingual interviewers for Spanish-speaking inmates, and by interviewer availability.

The response rates achieved on this unique effort were quite favorable. Of the 1,298 inmates selected, 1,173 (91 percent) completed the background questionnaire.<sup>1</sup> The assessment booklet was completed by 1,125 inmates and partially completed by another 22 inmates. These rates are a significant achievement, especially because the interviewers had no control over the availability of selected inmates within the short data collection period at each prison and the ability of inmates to complete the assessment. Although interview appointments were scheduled by the prison coordinator, sessions were sometimes delayed, interrupted, or canceled because of inmate transfers, unscheduled inmate count-downs, facility lockups, or movement of a sampled inmate to solitary confinement, restricted housing, or a unit for the mentally ill. Prisons also changed or canceled appointments to accommodate inmate obligations, such as attorney meetings or court appearances.

---

<sup>1</sup> This number includes 12 inmates who did not complete the background questionnaire for literacy-related reasons, including language problems and mental disabilities.

## **9.6 QUALITY CONTROL**

The measures used to ensure the collection of high-quality data included daily communications among interviewers, regional supervisors, and the home office. At the first two assigned facilities, each interviewer telephoned the home office sampling coordinator to review the sampling results immediately after completing inmate selection. Thereafter, interviewers called the coordinator if problems or questions arose during the sampling process. Any problems were referred to the data collection contractor's statistical staff before the sampling and interviews could proceed. Interviewers also reported the sampling results to their regional supervisor at the conclusion of the sampling process. Finally, after each day of interviewing, interviewers were required to contact their supervisor to discuss any special issues or concerns about the facility or inmate interview process.

Special editing specifications were prepared for the Prison Study. Editors were trained on the requirements and documents specific to the prison interviews.

*This page is intentionally left blank.*

## CHAPTER 10

### REDUCING THE RISK OF DATA DISCLOSURE

Thomas Krenzke, Sylvia Dohrmann, and Laura Alvarez-Rojas, Westat

Over the past decade, concerns about the disclosure of information related to individual survey respondents have increased. New laws have been put in place since the Privacy Act of 1974 to further ensure the protection of confidential data. The National Center for Education Statistics (NCES) and data contractors pledge confidentiality to respondents. The Education Sciences Reform Act of 2002 explicitly requires that NCES protect the confidentiality of all those responding to NCES-sponsored surveys so that no individual respondent can be identified. More specifically, NCES Standard 4-2, *Maintaining Confidentiality* (NCES 2002), provides guidelines for limiting the risk of data disclosure for data released by NCES. Data disclosure occurs when an individual respondent has been identified through the use of the survey item responses and other external data sources. This chapter describes the procedures for reducing the risk of data disclosure for the National Assessment of Adult Literacy (NAAL) in accordance with the guidelines specified in NCES Standard 4-2.

Several types of data were collected and derived during the NAAL sampling, data collection, and weighting processes. These variables were reviewed to determine their disclosure risk levels. The confidentiality analysis used a three-step process to reduce disclosure risk: (1) determining the disclosure risk arising from existing external data; (2) coarsening the data (described in section 10.1.2); and (3) swapping the data (described in section 10.3). Westat conducted the risk analysis, coarsening, and data swapping procedures to produce the following files containing data from all components of NAAL:

- Household Study public-use microdata file (PUMF);
- Prison Study PUMF;
- Household Study restricted-use microdata file (RUMF); and
- Prison Study RUMF.

Following the NCES guidelines, the RUMFs contain noncoarsened, swapped data, and the PUMFs contain coarsened and swapped data. Access to the RUMFs is restricted because users need a license to obtain the file. The PUMFs will be readily accessible to the public. The data swapping step for the RUMF was consistent with the one used for the PUMF to ensure consistency between statistics derived from the two datasets. In addition, the following confidentiality procedures were established for disseminating data through the RUMFs and PUMFs:

- identify personal identifiers, geographic information, and contextual variables (variables that can indirectly identify a geographic area);
- evaluate the existence of other publicly available files;
- evaluate the disclosure risk associated with release of the sampling and variance estimation variables; and
- evaluate the disclosure risk associated with release of key variables (i.e., visible variables) through extensive frequency tables.

Sections 10.1 and 10.2 discuss some outcomes of the risk analysis for the household sample and the prison sample, respectively. A general discussion of the data swapping procedures is provided in section 10.3.

## **10.1 HOUSEHOLD SAMPLE**

One aspect of the disclosure risk analysis for the PUMF was a review of each background questionnaire variable and groups of background questionnaire variables to determine whether any of the data presented a nonnegligible risk of individual disclosure. Several types of variables were available from the household sample and were analyzed for disclosure risk. These included variables collected through the survey and assessments, as well as variables created during weighting. These variables are summarized below:

- case identifiers;
- disposition codes for the survey instruments: screener, background questionnaire, assessment, and oral module;
- demographics: age, race/ethnicity, and gender;
- variance stratum and variance unit;
- sampling weights: weights from all stages of weighting, including the base weight, nonresponse-adjusted weight, trimmed weight, and final weight;
- weight adjustment factors, including compositing factors;
- number of eligible persons in the household and number selected;
- weighting variables, such as age category, census region, metropolitan statistical area identifier, educational attainment, race/ethnicity, and country of birth, and including imputation flags for educational attainment and country of birth;
- Background questionnaire data responses, including census codes for the respondent's industry and occupation;

- other background questionnaire variables, such as start and end times of the interview; and
- assessment scores.

Careful attention was given to these background questionnaire items and combinations of items. Even a very limited amount of demographic detail—such as income, occupation, age, year of immigration to the United States, foreign language spoken, and country of birth—can increase the chance that an individual can be identified. As discussed in section 10.1.1, personal and geographic identifiers were removed. Section 10.1.2 presents outcomes from the risk analysis in the form of variable suppression and recodes.

### **10.1.1 Personal Identifiers and Geographic Identifiers**

Any information that might be used to directly identify respondents and/or sampled locations was suppressed from the PUMF. This information included direct personal identifiers such as names (only first names were collected), addresses, and telephone numbers. Explicit geographic identifiers, such as state or county, were also suppressed, with the exception of census region and state identifiers for the State Assessment of Adult Literacy (SAAL). The SAAL states were Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma. The primary sampling unit (PSU) identification number (ID) was also suppressed. Further, because the case identifiers assigned during sample design, sample selection, and data collection had embedded geographic identifiers, the original case IDs were replaced with a sequential number to mask any pattern (e.g., alphabetic or geographic order).

### **10.1.2 Data Coarsening**

In general, data coarsening includes several types of procedures that decrease disclosure risk by reducing the amount of information released. Coarsening approaches include removing direct identifiers, limiting geographic detail, categorizing continuous variables, performing top- and bottom-coding,<sup>1</sup> and recoding values into broader categories. Weight adjustment factors and intermediate weights were removed because they posed a disclosure risk and provided minimal analytical value. Targeted or local suppression was also performed by removing the sensitive item value from the record or suppressing (or deleting) the variable from the file. During the NAAL data-coarsening step, some variables with high disclosure risk were suppressed or recoded; the original data values were retained for low-risk variables. The swapping procedure was used to add uncertainty to the otherwise individual identifying variables.

---

<sup>1</sup> With top-coding, the largest values of a variable are replaced with an upper limit, reducing the appearance of outlier data. Similarly, bottom-coding replaces the smallest values with a lower limit.

Variables with potential for high risk of disclosure were those known facts about individuals as collected by NAAL. These include income, race, occupation, and personal and geographic identifying variables. Variables with high disclosure risk that could not be recoded further were suppressed from the PUMF. The information contained in these variables was too specific to be released and could not be effectively recoded. Such variables include, for example, text responses to other-specify items. These types of items may have high risk of disclosure.

As described in the remainder of this section, seven major classes of variables were coarsened: age, race/ethnicity, language, education, income, occupation/industry, and all others. The recodes are a result of many cross-tabulations between background questionnaire variables and demographic/geographic variables. Of primary concern were cases that were rare in the population. Data swapping was also implemented to further protect the identity of the individual.

**Age.** The respondent's date of birth, collected on the background questionnaire, was converted to a single year of age. Whenever age was missing, it was reconciled with the screener data (the screener asked for a single year of age for each person in the household). The screener had a backup question that requested the response in age categories. These screener age categories were used to impute a single year of age for respondents still missing an age; the single year of age was released on the RUMF. The single years were categorized as shown in table 10-1 for the PUMF. Other age-related recoded variables are also shown in table 10-1.

**Table 10-1. Source, description, and categories of age-related recoded variables for household sample: 2003**

| Variable name | Source variables       | Variable description  | Categories of recoded variable  |
|---------------|------------------------|---|---|
| DARRIVE       | BQ1029                 | Immigration age   | U.S. born; 0–18; 19 or more years   |
| DLIVEUS       | BQ1030                 | Years lived in United States  | Combined highest age categories: 1–5; 6 or more years   |
| DENGAGE       | BQ1065                 | Age when respondent spoke English   | 1–10; 11 or more years; does not speak English  |
| DAGE          | CALCAGE                | Age calculated from date of birth   | Included screener data (where background questionnaire item was missing) to recode into categories: 16–18; 19–24; 25–39; 40–49; 50–64; 65 or more years |
| DHSAGE        | BQ1032, BQ1205, BQ1210 | Age when respondent graduated from high school or obtained General Educational Development (GED) credential | 16–19; 20 or more years; not applicable   |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

**Race/ethnicity.** Table 10-2 contains the race/ethnicity variables and the recodes. The recode for respondent race/ethnicity used screener data when background questionnaire items were missing.

**Table 10-2. Source, description, and categories of race/ethnicity recoded variables for household sample: 2003**

| Variable name | Source variables               | Variable description                       | Categories of recoded variable                        |
|---------------|--------------------------------|--|---|
| DCBIRTH       | BQ1025, BQ1025S                | Country where respondent was born          | United States; other                                  |
| DMCBIRTH      | BQ1875, BQ1875S                | Country where respondent's mother was born | United States; other                                  |
| DFCBIRTH      | BQ1885, BQ1885S                | Country where respondent's father was born | United States; other                                  |
| DRACE         | BQ2440, BQ2445_a-e, BQ2450_a-e | Race/ethnicity                             | White, Black, Hispanic, other (including multiracial) |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

**Language.** Table 10-3 shows the language variables and the method of coarsening. All items that allow several responses, including other-specify variables (i.e., variables containing text responses not contained in the response categories provided), were coded into five categories:

- English only;
- English and Spanish (with or without other);
- English and other;
- Spanish only or with other; and
- other only.

This coding scheme preserved the most frequent responses (English and Spanish) while capturing multilingual respondents and those who spoke no English. Recoding for the variable indicating other languages the respondent speaks now includes the language he or she speaks best.

**Table 10-3. Source, description, and categories of language recoded variables for household sample: 2003**

| Variable name | Source variables         | Variable description                                | Categories of recoded variable   |
|---------------|--------------------------|---|--|
| DHMLANG       | BQ1045a-e,<br>BQ1045Sa-e | Language spoken at home                             | Five categories: English only; English and Spanish (with or without other); English and other; Spanish only or with other; other only                                      |
| D1STLAN       | BQ1050a-e,<br>BQ1050Sa-e | Language spoken before school                       | Five categories: English only; English and Spanish (with or without other); English and other; Spanish only or with othe; other only                                       |
| DLANGRW       | BQ1060,<br>BQ1060S       | Language respondent first learned to read and write | English; Spanish; other  |
| DCLANGS       | BQ1090,<br>BQ1090S       | Language respondent usually speaks now              | English; Spanish; other  |
| DOLSOPT       | BQ1095a-e,<br>BQ1095Sa-e | Other language often spoken now                     | Incorporated BQ1090 and recoded into five categories: English only; English and Spanish (with or without other); English and other; Spanish only or with other; other only |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

**Education.** Table 10-4 shows education variables and the categories of their recodes. The recode for educational attainment consists of nine categories:

- still in high school;
- less than high school/some high school;
- General Educational Development (GED)/high school equivalency;
- high school graduate;
- vocational/trade/business school;
- some college;
- associate's/2-year degree;
- college graduate; and
- graduate studies/degree.

**Table 10-4. Source, description, and categories of education recoded variables for household sample: 2003**

| Variable name | Source variable | Variable description                         | Categories of recoded variable   |
|---------------|-----------------|--|--|
| DEDBFUS       | BQ1040          | Education before coming to the United States | Did not attend school/primary, elementary, secondary or more   |
| DEDATTN       | BQ1205          | Highest education level attained             | Still in high school (HS); less than HS/some HS; General Educational Development/HS equivalency; HS graduate; vocational; some college; associate's degree; B.S. or B.A.; graduate studies/degree  |
| DDTYPE        | BQ1225          | Diploma type                                 | Regular diploma from school in United States or U.S. government school outside United States; regular diploma from school outside United States; General Educational Development credential or certificate of completion; combined "Did not receive HS diploma" with missing |

See note at end of table.

**Table 10-4. Source, description, and categories of education recoded variables for household sample: 2003—Continued**

| Variable name | Source variable | Variable description                              | Categories of recoded variable   |
|---------------|-----------------|---|--|
| DRFSSCHC      | BQ1235          | Reason respondent stopped school                  | Financial problems; did not do well in school; did not like school or was bored in school; was expelled from school or asked to leave; wanted to work; wanted to go into the military; personal reasons; other |
| DSGRDHS       | BQ1245          | State in which respondent graduated high school   | Current; other   |
| DSGRDCO       | BQ1257          | State in which respondent received college degree | Current; other   |
| DMED          | BQ1880          | Mother's highest educational attainment           | Less than HS/some HS; General Educational Development credential/HS equivalency; HS graduate; vocational; some college; associate's degree; B.S. or B.A.; graduate studies/degree                              |
| DFED          | BQ1890          | Father's highest educational attainment           | Less than HS/some HS; General Educational Development credential/HS equivalency; HS graduate; vocational; some college; associate's degree; B.S. or B.A.; graduate studies/degree                              |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

**Income.** The household background questionnaire included several income-related questions; these included some weekly wage amounts, as well as income amounts for the entire year. Using the variables in table 10-5, variables were derived for the weekly wage during the previous week, income adequacy, personal income, and household income.

Data on total personal and household income were gathered through a series of questions. First, the respondent was asked one question with 13 categories (14 in the case of household income). If the respondent refused that question, he or she was asked a series of questions designed to categorize income into one of eight categories. The income variable was created in the eight categories to use all of the information possible.

**Table 10-5. Source, description, and categories of income-related recoded variables for household sample: 2003**

| Variable name | Source variables  | Variable description                        | Categories of recoded variable   |
|---------------|---|---|--|
| D_WKLYWAGE    | BQ1500,<br>BQ1505,<br>BQ1505S,<br>BQ1515,<br>BQ1520                                 | Weekly wage (previous week)                 | Less than \$300;<br>\$300–\$499;<br>\$500–\$649;<br>\$650–\$1,149;<br>\$1,150–\$1,949;<br>greater than or equal to \$1,950   |
| DINCOME       | BQ1895,<br>BQ1910a-j,<br>BQ1920a-j,<br>BQ2430                                       | Income adequacy                             | Below poverty threshold; other   |
| DBQ2421       | BQ2421,<br>BQ2422,<br>BQ2423,<br>BQ2424,<br>BQ2425,<br>BQ2426,<br>BQ2427,<br>BQ2428 | Approximate personal income                 | Less than \$5,000;<br>\$5,000–\$9,999;<br>\$10,000–\$14,999;<br>\$15,000–\$19,999;<br>\$20,000–\$29,999;<br>\$30,000–\$39,999;<br>\$40,000–\$59,999;<br>\$60,000 or more     |
| DBQ2430       | BQ2430,<br>BQ2432,<br>BQ2433,<br>BQ2434,<br>BQ2435,<br>BQ2436,<br>BQ2437,<br>BQ2438 | Approximate household income                | Less than \$10,000;<br>\$10,000–\$14,999;<br>\$15,000–\$19,999;<br>\$20,000–\$29,999;<br>\$30,000–\$39,999;<br>\$40,000–\$59,999;<br>\$60,000–\$99,999;<br>\$100,000 or more |
| DWFTIME       | BQ2155  | Length of participation in welfare programs | Less than 2 years; 2 or more years   |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

**Occupation and industry codes.** Each respondent was asked to specify his or her occupation and industry. These responses were then categorized into the standard four-digit values used by the U.S. census. These detailed codes had a high disclosure risk. However, the census also uses some standard combinations of two-digit classifications (such as combining all codes beginning with 01 with all codes beginning with 02). Using these classifications reduced the risk of data disclosure.

Other derived variables are provided in appendix A.

## 10.2 PRISON SAMPLE

As for the household sample, the statistical disclosure control procedures for the prison sample were based on a disclosure risk analysis. After the sources of disclosure risks had been identified, variable suppression, coarsening, and swapping procedures were applied to reduce such risks. The application of these techniques followed NCES guidelines, specifically NCES Standard 4-2, *Maintaining Confidentiality* (NCES 2002).

All variables collected or derived through the sampling, data collection, and weighting process were compiled. The following list shows the variables available and analyzed for data disclosure risk:

- case identifiers;
- disposition codes for the background questionnaire, assessment, and oral module;
- variance stratum and variance unit;
- prison-level variables: census region (Northeast, Midwest, South, West); type of prison (federal, state, private); gender composition of the prison (male only, female only, mixed); and security type (supermaximum, maximum, medium, minimum, administrative, other);
- inmate full sample weight and replicate weights;
- adjustment factors and intermediate weights (prison base weight, prison nonresponse-adjusted weight, inmate base weight); and
- weighting variables, such as imputation flags for weighting purposes and raking dimensions.

The background questionnaire collected the following data responses:

- Section A: General and Language Background;
- Section B: Educational Background and Experiences;
- Section C: Prison Experiences;
- Section D: Prison Work Assignments and Labor Force Participation;
- Section E: Political and Social Participation;
- Section F: Literacy Practices;
- Section G: Demographic Information;
- Section H: Household Income and Welfare Participation;

- Section I: Health Questions; and
- Section J: Additional Demographics (race/ethnicity variables).

After the variables had been compiled, personal identifiers and geographic information were treated as discussed in section 10.2.1. Weight adjustment factors and intermediate weights were removed because they posed a disclosure risk and provided minimal analytical value. The ability to match with administrative data and other external data was assessed, as discussed in section 10.2.2. Data coarsening procedures are discussed in section 10.2.3, and section 10.3 provides a general discussion of data swapping.

### **10.2.1 Personal Identifiers and Geographic Information**

Any information that could be used to directly identify persons and/or prisons was suppressed from the file. This information includes the following:

- Direct and proximate identifiers:
  - direct identifiers of prisons such as names, addresses, and telephone numbers and names of inmates.
- Explicit geographic information:
  - Explicit geographic identifiers (except census region).

Further, because the original numerical IDs provide key identifying information, the case IDs were replaced by a sequential number to mask any pattern, such as alphabetic or geographic order.

### **10.2.2 Ability to Match With Administrative Data and Other External Data**

The Bureau of Justice Statistics 2000 Census of State and Federal Adult Correctional Facilities (referred to as the Census) is a prison-level data source for which data are publicly available and which presents a potential risk of disclosure. The Census includes more than 1,600 facilities. The Census data include address, capacity, inmate population, and security level, all of which are important characteristics for sampling and data collection. A record linkage analysis was conducted but did not show a risk of disclosure from matching to the Census data. “Record linkage analysis” is a technical term relatively well known among researchers of disclosure risk. Skinner and Elliott (2002) provide a good discussion of record linkage techniques.

### 10.2.3 Data Coarsening

The disclosure analysis for the PUMF included a review of each background questionnaire variable to determine whether any of the proposed data presented a nonnegligible risk of individual disclosure. For the disclosure risk analysis, frequencies were processed. If the number of observations for a category was low, then the variables were generally recoded into broader classifications or were suppressed. Once the initial coarsening process was complete, cross-tabulations between key reporting variables and several identifiable variables (education, race/ethnicity, age, and other background questionnaire items) were processed. The cross-tabulations further determined the risk level of each variable. The analysis of the cross-tabulations identified more recodes, and decisions were made about the inclusion of certain high-risk variables on the file. Concerns about some of the variables with a high risk of disclosure were reduced through variable suppression. For these variables, the risk of disclosure could not be reduced further by recoding the data.

As shown in table 10-6, year of immigration to the United States, the respondent's age, and the language variables are among the variables that were reclassified. The industry and occupation codes proposed for the household sample were used for the prison sample.

**Table 10-6. Source, description, and categories of recoded variables for prison sample public-use microdata file: 2003**

| Variable name | Source variables | Variable description                                | Categories of recoded variables                                 |
|---------------|------------------|---|---|
| DAGEC         | CALCAGE          | Age calculated from date of birth                   | Age: 16–24; 25–39; 40–49; 50 or more years                      |
| DCBIRTH       | BQ1015           | Country of birth                                    | United States (50 states and DC); other                         |
| DARRIVE       | BQ1020           | Immigration age                                     | U.S. born; 0–18; 19 or more years                               |
| DEDBFUS       | BQ1050           | Education before coming to United States            | Did not attend school/primary, elementary, secondary or more    |
| DHMLANGC      | BQ1055a-e        | Language spoken at home when growing up             | English only; English and other (including Spanish); other only |
| D1STLANG      | BQ1060a-e        | Language spoken before school                       | English only; English and other (including Spanish); other only |
| DLANGRWC      | BQ1070           | Language respondent first learned to read and write | English; other  |

See note at end of table.

**Table 10-6. Source, description, and categories of recoded variables for prison sample public-use microdata file: 2003—Continued**

| Variable name | Source variables       | Variable description                                  | Categories of recoded variables   |
|---------------|------------------------|---|---|
| DENGAGE       | BQ1075                 | Age respondent learned to speak English               | 1–10; 11 or more years; do not speak English  |
| DCLANGSC      | BQ1100, BQ1105a-c      | Language respondent usually speaks now                | English, other; English only; English and other; other only   |
| DOLANGSB      | BQ1110                 | Other language respondent speaks best                 | Spanish; other  |
| DEDATTNC      | BQ1205, BQ1208, BQ1215 | Educational attainment                                | Less than high school (HS)/some HS; General Educational Development credential or HS equivalency; HS graduate; vocational, trade, or business school after HS; less than 2 years of college; associate’s degree; college or more                  |
| DHSAGE        | BQ1030, BQ1205, BQ1220 | Age upon graduating from high school                  | 16–19; 20 or more years; not applicable   |
| DDTYPEC       | BA1224                 | Type of high school degree                            | Regular diploma from school in United States, or U.S. government school outside United States; General Educational Development credential or certificate of completion; “Did not receive HS diploma”; “Regular from school outside U.S.”; missing |
| DRFSSHC       | BQ1230                 | Reason for stopping school before college degree      | Financial problems; did not do well in school; did not like school or was bored in school; expelled from school or asked to leave; wanted to work; wanted to go into the military; personal reasons; sent to jail or detention or prison; other   |
| DPVOC         | BQ1421, BQ1423         | Length of time in prison Vocational training programs | Less than 1 year; 1 or more years   |
| DPCLSHR       | BQ1450                 | How many hours spent in prison classes last week      | 0, 1–19; 20–49; 50 or more hours  |
| DOFFENS1-3    | BQ1475a-c              | Offenses for which respondent is in prison            | Violent; property; drugs; public order; other   |

See note at end of table.

**Table 10-6. Source, description, and categories of recoded variables for prison sample public-use microdata file: 2003—Continued**

| Variable name | Source variables     | Variable description             | Categories of recoded variable   |
|---------------|----------------------|----------------------------------|--|
| DEMPTYPC      | BQ1620               | Type of employer in past 3 years | Private employer; self-employed; other   |
| MEDC          | BQ1830               | Mother's educational attainment  | Less than HS/some HS; General Educational Development credential or HS equivalency; HS graduate; vocational, trade, or business school after HS; less than 2 years of college; associate's degree; college or more |
| FEDC          | BQ1840               | Father's educational attainment  | Less than HS/some HS; General Educational Development credential or HS equivalency; HS graduate; vocational, trade, or business school after HS; less than 2 years of college; associate's degree; college or more |
| DMCBIRTH      | BQ1825               | Mother's country of birth        | United States; other   |
| DFCBIRTH      | BQ1835               | Father's country of birth        | United States; other   |
| DMARITAL      | BQ1845               | Marital status                   | Never married; married or living as married; separated or divorced or widowed  |
| DWLFLSTC      | BQ1880               | Last received welfare payments   | 3 years or less; more than 3 years   |
| DRACE         | BQ2000,<br>BQ2010a-e | Race/ethnicity                   | White; Black; Hispanic; other (including multiracial)  |

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

### 10.3 DATA SWAPPING

To ensure that an individual respondent is not identified, the risk of data disclosure was further reduced by using a data swapping technique. Data swapping is an NCES requirement that reduces risk by modifying microdata. In data swapping, a probability sampling of records are paired with other records on the file using selected characteristics, and then some identifying variables are swapped between the two records (refer to Kaufman et al. [2005] for further discussion). The sampling rate for NAAL swapping was designed to protect the confidentiality of the data without affecting the usability of the dataset. This

method is an effective way of keeping as much valuable data as possible while protecting respondent identity.

Swapping preserves the univariate frequencies, means, and variances, although it may affect multivariate relationships. Pre- and post-swapping percentage distributions (unweighted and weighted) and correlations were reviewed to ensure data quality was maintained.

*This page is intentionally left blank.*