

## CHAPTER 4

### FIELD-TEST SCORING AND ANALYSIS OF FIELD-TEST DATA

Elizabeth Greenberg, American Institutes for Research

Following the conclusion of the field test, the field-tested cognitive items were scored and the results were analyzed to determine which items to retain for the operational assessment. The background questionnaire (BQ) data obtained during the field test were analyzed, and changes were made to the BQ on the basis of the field-test data. The field-test results were also used to select the core items for the operational assessment and to develop the algorithm for selecting Adult Literacy Supplemental Assessment (ALSA) respondents.

#### 4.1 DEVELOPMENT OF SCORING PROCEDURES

This section discusses the refinement of the scoring rubrics on the basis of range-finding conducted using the field-test booklets, the development of scorer training materials, and the field-test scoring procedures.

##### 4.1.1 Refinement of Scoring Rubrics

Draft scoring rubrics for the items in the National Assessment of Adult Literacy (NAAL) field test were developed by the item writers. Prior to the scoring of the NAAL field test, project staff conducted range-finding—reviewing a sample of assessment booklets to determine the range of responses received for a particular item—to identify the different types of responses received for the individual NAAL items. Approximately 100 responses to each item were reviewed as part of the range-finding process, and the range-finding staff photocopied all distinct responses to each item to use during scorer training. The scoring rubrics were refined and expanded to address alternative correct or partially correct responses identified during range-finding. For example, one assessment item asked respondents to contrast two different ideas. Although scorers will always need to exercise some judgment when evaluating the responses to this type of question, range-finding identified the most common responses. These responses were then summarized on the scoring rubrics, with the proper score assigned to each response.

#### **4.1.2 Development of Scorer Training Materials**

The responses collected during range-finding were used as the basis for the scorer training materials. Responses that closely matched each score point on the scoring rubric were identified during range-finding for use as anchor papers to illustrate each score point. Both straightforward and ambiguous responses for each score point were identified as training papers for the scorers to score individually and then discuss as a group. The number of responses used as training papers varied by item. For straightforward items, which required a number or short phrase as a response, five or six training papers were identified. For items for which the responses were longer and more complex to score, up to 20 training papers were identified.

Project staff developed a detailed script for each item for the scorer trainers to follow. The script for each item began by asking the scorers to read the item and formulate an answer. Next, the script walked the scorers through the rubric, explaining what a respondent had to do to receive credit for an item. Where applicable, the script instructed the trainers to remind the scorers to be flexible when interpreting responses and to allow variations in wording that still expressed the meaning of the original score point.

Next, the script directed the trainers to ask whether the scorers had any questions and then to walk the scorers through the anchor papers. The script included a table that listed each anchor paper and explained why it was chosen. After going through the anchor papers, the scorers were directed by the trainers to individually score the training papers for a particular item. Another table in the script listed the correct score point for each training paper and an explanation of why that score point was assigned. The script directed the trainers to discuss each training paper with the scorers and to not move on until all scorers indicated that they understood why each paper was assigned a particular score.

#### **4.1.3 Field-Test Scoring Procedures**

Field-test scoring was conducted in Iowa City, Iowa. Because the NAAL field-test scoring took place during January, which is normally a slow time for testing organizations, NAAL was able to use only scorers who had significant experience scoring open-ended language arts items. Twenty scorers were hired for the NAAL field test. They were divided into two groups, called tables, with nine scorers and a table leader at each table.

Item development staff traveled to Iowa to train the scorers. Each table was trained on one block at a time and completed the scoring of that block before moving on to score the next block. Because this

was a field test and the first time these scoring rubrics and scorer training materials were used, one purpose of the scoring was to further refine both the rubrics and the training materials. Item development staff discussed the rubrics and the training materials with the scoring staff as they were scoring and noted the training papers for which more detailed explanations of the assigned scores would be helpful. They also noted the items for which additional training papers or more detailed rubrics were needed to address ambiguities. In a few cases, rubrics were edited during the field-test scoring process.

## **4.2 ANALYSIS OF FIELD-TEST DATA**

This section describes the analysis of the NAAL field-test data, including background questionnaire responses, field-test cognitive data, the Fluency Addition to the NAAL (FAN), and the Adult Literacy Supplemental Data (ALSA).

### **4.2.1 Background Questionnaire Analyses**

Project staff analyzed the NAAL field-test BQ data to identify

- questions that respondents refused to answer;
- questions for which the relationship to the respondent's literacy level was different than expected;
- questions with unexpected response patterns (e.g., large numbers of respondents picked a response that goes against known research);
- redundant questions (i.e., questions measuring the same construct with high response correlations);
- problems with skip patterns;
- places where look-up tables could replace open-ended questions to make the analysis easier; and
- any difficulties with the administration of the BQ in the field.

All field interviewers were asked to complete a questionnaire identifying any problems they encountered when administering the BQ in the field, and project staff then reviewed their responses. Additionally, approximately half a dozen field interviewers were asked to participate in a debriefing session to discuss any problems they encountered when administering the BQ in the field. Field staff all reported that the BQ was easy or very easy to administer.

Project staff computed frequencies for every question on the BQ and mean booklet scores for all response categories for each background question. Project staff also ran factor analyses for sets of questions that were designed to measure different aspects of the same underlying construct.

Almost all the questions on the BQ functioned as expected. As a result of the factor analyses, three questions that overlapped with other questions and were not necessary to measure an underlying construct were dropped. A question asking respondents to identify the city and county they lived in when they graduated high school was dropped because the results could not be matched to existing databases by using Federal Information Processing Standards (FIPS) codes or ZIP codes; however, the question asking respondents which state they lived in when they graduated high school was retained. Skips were added in a few places where questions were not relevant to some of the respondents, and one problem with the computer-assisted personal interviewing (CAPI) programming for a skip pattern was identified and corrected. Reference time periods were changed for a couple of questions for which interviewers indicated that respondents were confused or unable to answer a question.

Administration time for the field-test BQ averaged 26 minutes, which was 9 minutes less than the time allocated. This provided some flexibility, and five questions were added to the BQ, including one asking respondents when they completed their last year of college and one asking about veteran status.

#### **4.2.2 Analysis of Field-Test Cognitive Data**

The field test consisted of 12 blocks of items,<sup>1</sup> with a total of 141 noncore items, plus four versions of the core, with a total of 17 core items (see section 2.1 for a discussion of core and noncore items). The noncore blocks were assembled in a partial spiral so that each block appeared in each position in a test booklet, but not every block appeared with every other block. Table 4-1 shows the field-test booklet spiral.

---

<sup>1</sup>The 12 non-core blocks were numbered 1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 4.1, 4.2, and 4.3.

**Table 4-1. Field-test booklet spiral: 2003**

Booklet #	Position 1	Position 2	Position 3	Position 4
1	Core 1	1.1	1.2	1.3
2	Core 2	2.1	2.2	2.3
3	Core 3	3.1	3.2	3.3
4	Core 4	4.1	4.2	4.3
5	Core 1	1.2	1.3	1.1
6	Core 2	2.2	2.3	2.1
7	Core 3	3.2	3.3	3.1
8	Core 4	4.2	4.3	4.1
9	Core 1	1.3	1.1	1.2
10	Core 2	2.3	2.1	2.2
11	Core 3	3.3	3.1	3.2
12	Core 4	4.3	4.1	4.2

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

Project staff analyzed the NAAL field-test cognitive data to identify

- items with differential item functioning (DIF) in regard to female versus male, Hispanic versus other, Black versus other, or age 60+ versus under age 60;
- items with interrater reliability problems;
- items that did not discriminate well among respondents with different levels of literacy;
- noncore items with field-test p-values (percentage of respondents who answered each item correctly) outside the range of the 1992 noncore items that were being replaced (below .20 or above .90);
- items appropriate for partial credit; and
- a set of core items that could be used to screen respondents for inclusion in ALSA or the main NAAL.

Because the field test was designed to oversample Blacks and Hispanics, weights were created prior to the field-test data analysis so that the percentages of Blacks, Hispanics, and people age 60 or older corresponded to the percentages in the 2000 Census.

### 4.2.2.1 Item Analyses

As shown in table 4-1, booklets 1, 5, and 9 included the same blocks and items. Similarly, booklets 2, 6, and 10 included the same blocks; booklets 3, 7, and 11 included the same blocks; and booklets 4, 8, and 12 included the same blocks. Item analysis was done separately for each of the four sets of field-test booklets, using weighted data. Because of the nature of the field-test sample, which combined respondents selected through a probability sample with a convenience sample of respondents selected through focus group centers (see chapter 3), the field-test data could not be weighted to accurately reflect population estimates. However, the data were weighted to roughly reflect the racial/ethnic breakdown of the population (Black, Hispanic, other) and the age breakdown of the population based on the 2000 Census.

Project staff computed the following statistics for each item:

- percentage of respondents who answered the item correctly ( $p$ -value);
- mean booklet score for each score point associated with an item;
- differential item functioning (DIF; female vs. male, Hispanic vs. other, Black vs. other, or age 60+ vs. under age 60);
- biserial correlation; and
- interrater reliability.

Each of these statistics is discussed below.

#### ***Percentage of Respondents Who Answered Each Item Correctly ( $p$ -value)***

The size of the field test necessitated obtaining data from each respondent for all the items with which the respondent was presented. Therefore, field-test blocks were purposely kept short so that respondents could easily complete all the items in a block. Because of this decision, project staff did not distinguish between “not reached” and “missing/skipped” when analyzing the field-test data and treated all blank responses as “incorrect” when calculating  $p$ -values for the field-test items. Noncore items with field-test  $p$ -values outside the range of the 1992 items that were being replaced (below .20 or above .90) were flagged. Of the 141 noncore items in the field test, 3 had  $p$ -values below .20 and 3 had  $p$ -values above .90 (table 4-2).

**Table 4-2. Number of NAAL noncore and core field-test items, by percent correct (p-value): 2003**

<i>p</i> -value	Number of noncore items	Number of core items
Below .20	3	0
.20 to .90	135	8
Above .90	3	9

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

### ***Mean Booklet Score***

The mean booklet score for a score point is the average number of items that respondents choosing that score point answer correctly in the remainder of the booklet in which the item of interest appears. Generally, respondents who receive a score of correct on an item should have a higher mean booklet score than respondents who do not answer the item correctly. If they do not, this is an indication that the item is not functioning as expected. Mean booklet scores for each score point—including “any other response,” blank, off task, illegible, and “I don’t know”—were computed for each item. No items had incorrect responses that had mean booklet scores equal to or higher than the correct response.

### ***Differential Item Functioning***

Differential item functioning (DIF) analysis refers to procedures that assess whether specific items are differentially difficult for different groups of respondents after controlling for overall differences among groups. DIF procedures compare the performance of groups on each item within sets of respondents who have the same level of performance, usually measured by total test score. Items identified as having DIF are then evaluated to determine whether they are biased; that is, whether the DIF is related to a factor unrelated to what is being tested. A biased item is generally deleted from a test because the probability of doing well on the item depends in part on characteristics of the item that are not related to the construct being measured.<sup>2</sup>

Items were classified as A, B, or C in regard to DIF:

- A means the item has negligible DIF.
- B means the item has moderate DIF.

---

<sup>2</sup> This paragraph was adapted from the forthcoming National Assessment of Educational Progress (NAEP) Technical Report.

- C means the item has significant DIF.

DIF classifications were based on delta ( $\hat{\Delta}_{MH}$ ) computed as follows:

$$\hat{\Delta}_{MH} = -\frac{4}{1.7} \ln(\hat{\alpha}_{MH}) = -2.35 \ln(\hat{\alpha}_{MH})$$

where  $\hat{\alpha}_{MH}$  is the Mantel-Haenszel (MH)-odds ratio.<sup>3</sup> Statistical tests of the null hypothesis were performed at the .05 level using the Chi-square test statistics MH (chi-square), which is distributed as a chi-square with 1 degree of freedom (table 4-3).

**Table 4-3. Classification rules for differential item functioning (DIF) categories: 2003**

Category	Classification rule
C	$ \hat{\Delta}_{MH} $ is significantly greater than 1.0, and $ \hat{\Delta}_{MH}  > 1.5$ .
B	$ \hat{\Delta}_{MH} $ is significantly different from zero and $\geq 1.0$ , and either a) $ \hat{\Delta}_{MH}  \leq 1.5$ or b) $ \hat{\Delta}_{MH} $ is not significantly greater than 1.0
A	$ \hat{\Delta}_{MH} $ is not significantly different from zero using $MH\chi^2$ , or $ \hat{\Delta}_{MH}  < 1.0$

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

DIF was classified either + or – depending on whether the items favored the target group or the reference group. Items that favored the target group were classified +, and items that favored the reference group were classified –. DIF analysis was conducted for the following target groups: females versus males, Hispanics versus others (others included all adults who were not classified as Hispanic or Black), Blacks versus others (others included all adults who were not classified as Hispanic or Black), and adults age 60 or over versus under age 60. Items with B or C DIF were flagged for further examination. Table 4-4 shows the number of items that exhibited DIF in the field test.

<sup>3</sup> For a discussion of delta, see Holland, P.W., and Thayer, D.T. (1988). *Differential Item Performance and Mantel-Haenszel* in Wainer, H. & Braun, H. (Eds.). *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

**Table 4-4. Number of NAAL noncore and core field-test items, by differential item functioning (DIF) categories: 2003**

Differential item functioning	Number of noncore items	Number of core items
Female vs. male: A	128	17
Female vs. male: B+	6	0
Female vs. male: B-	5	0
Female vs. male: C+	0	0
Female vs. male: C-	2	0
Hispanic vs. other: A	131	14
Hispanic vs. other: B+	2	2
Hispanic vs. other: B-	7	1
Hispanic vs. other: C+	0	0
Hispanic vs. other: C-	1	0
Black vs. other: A	132	17
Black vs. other: B+	5	0
Black vs. other: B-	4	0
Black vs. other: C+	0	0
Black vs. other: C-	0	0
60+ vs. under 60: A	123	13
60+ vs. under 60: B+	11	2
60+ vs. under 60: B-	4	2
60+ vs. under 60: C+	3	0

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

### ***Biserial Correlation***

The biserial correlation is an item discrimination index that is based on the weighted correlation between the item score and the booklet score, where the booklet score excludes the item being studied. The biserial correlation indicates the extent to which each item differentiates among those respondents who possess the skills being measured and those who do not. The higher the value of the biserial correlation, the more discriminating an item is. In general, a biserial correlation below .2 indicates an item problem. None of the field-test items was flagged because of issues related to biserial correlation. Among the field-test items, 136 had biserial correlations of .60 or higher, 21 had biserial correlations of .40 to .59, and 1 had a biserial correlation of .20 to .39 (table 4-5).

**Table 4-5. Number of NAAL noncore and core field-test items, by biserial correlation: 2003**

Biserial correlation	Number of noncore items	Number of core items
Above .60	124	12
.40 to .59	17	4
.20 to .39	0	1
Below .20	0	0

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

### ***Interrater Reliability***

Half the field-test booklets were scored by a second scorer, and interrater reliability statistics, comparing the score assigned by the first scorer with the score assigned by the second scorer, were computed for each field-test item. In addition to computing exact agreement between the first and second scorers, project staff computed the percentage of cases for which one scorer assigned full credit and the second scorer assigned partial credit, or one scorer assigned partial credit and the second scorer assigned no credit. Items with exact interrater reliability below .95 were flagged for further examination. For the field test, 140 items had interrater reliability of 95 percent or higher, 10 items had interrater reliability of 90 to 94.9 percent, 6 items had interrater reliability of 85 to 89.9 percent, 1 item had interrater reliability of 80 to 84.9 percent, and 1 item had interrater reliability below 80 percent (table 4-6).

**Table 4-6. Number of NAAL noncore and core field-test items, by interrater reliability (exact agreement): 2003**

Interrater reliability	Number of noncore items	Number of core items
95 or higher	125	15
90 to 94.9	8	2
85 to 89.9	6	0
80 to 84.9	1	0
Below 80	1	0

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

### **4.2.2.2 Flagged Items**

Items that were flagged during the item analysis of the field-test data were examined by senior project staff. All noncore items with *p*-values outside the range of .20 to .90 and all noncore items with biserial correlations below .60 were eliminated from the item pool. Core items with *p*-values above .90 were not considered to be a problem because the core items were deliberately kept easy to ease

participants into the main assessment blocks. Two core items with biserial correlations between .40 and .59 were also retained for the operational assessment.

Senior project staff also examined all items with scoring interrater reliability below 95 percent. The single item with interrater reliability below 80 percent was the only item on the field test that asked the respondent to give an oral response that was recorded for scoring later. That item was eliminated because of the difficulty in scoring the oral response. The other items with interrater reliability below 95 percent were among the more-difficult-to-score items, but it seemed likely that interrater reliability could be improved for these items with additional scorer training. These items were not automatically eliminated from the item pool, but—as discussed in chapter 13—additional range-finding was conducted for these items and additional scorer training materials were developed for the seven items with interrater reliability below .95 that were retained for the operational assessment.

All items with C DIF were eliminated from the pool for assembling the operational blocks, with the exception of one item that had C+ DIF in favor of respondents age 60 or over. This item was kept because the stimulus material and question were determined to be important to the assessment. The review panel that evaluated the operational blocks (see chapter 2) was asked to pay particular attention to this item, and the panel endorsed the decision to keep this item in the assessment.

B DIF was considered less of a problem than C DIF, and items with B DIF were not automatically eliminated from the item pool. In general, these items were balanced in terms of whether they favored the target group or the reference group. Among the items in the seven new blocks assembled from the field-test items, 3 items exhibited female versus male B DIF (1 in favor of females, 2 in favor of males), 3 items exhibited Hispanic versus others B DIF (all in favor of others), 3 items exhibited Black versus others B DIF (1 in favor of Blacks, 2 in favor of others), and 11 items exhibited 60+ versus under 60 DIF (8 in favor of 60+, 3 in favor of under 60). No items retained for the core exhibited any DIF.

#### **4.2.2.3 Partial Credit**

For some of the NAAL field-test items, data were collected for potential partial-credit score points. For these items, the NAAL field-test data analysis was used to make preliminary decisions about whether the items supported awarding partial credit. These decisions were based on the weighted mean field-test booklet scores associated with the partial-credit score points. If the mean field-test booklet score associated with a partial-credit score point was between the mean booklet score associated with a correct response and a mean booklet score associated with an incorrect response, the partial-credit score point was retained for the operational scoring rubrics. For items for which the mean booklet score for the

partial-credit score point did not fall between the mean booklet score for full credit and the mean booklet score for no credit, the partial-credit score point was dropped from the scoring rubrics for the operational assessment.

#### **4.2.2.4 Algorithm Used to Identify Adult Literacy Supplemental Assessment Respondents**

One goal of the field-test data analysis was to select a set of questions that could accurately identify the bottom 10 percent of respondents as appropriate for the ALSA. The bottom 10 percent of respondents were targeted because on the basis of 1992 data, it was thought that they would be unlikely to complete enough NAAL items to provide meaningful data.<sup>4</sup> The analysis plan called for starting with one of the four versions of the core that was field-tested in fall 2001 and supplementing it as necessary with the fewest number of additional items that would accurately identify ALSA respondents. The additional items were chosen because of their ease of scoring (they needed to be accurately scored by field interviewers) and performance in the field test (items with DIF and items that did not discriminate well among respondents with different levels of literacy were eliminated from consideration). Because of DIF problems, the version of the core that appeared in booklets 1 and 3 was eliminated from consideration as the basis of the core for the operational assessment.

On the basis of the analysis of the field-test data, the core that appeared in field-test booklet 2, supplemented with question 9 from block 1 of that booklet, was identified as having the best combination of items to identify respondents who would likely fall in the bottom 10 percent (unweighted) of the population because of literacy—the screening criteria for the ALSA as discussed above. Question 9 in block 1 was a health question based on the *Medicare and You* booklet. The question directed respondents to a particular page in the booklet and asked them to locate information about how often people should get a flu shot.

In the NAAL field test, 350 respondents answered booklets 2, 6, or 10 (as shown in table 4-1, these booklets included the same set of cognitive items, with the three blocks presented in a different location in each book). A rank ordering of these respondents from lowest to highest score showed that the 35th respondent correctly answered 4.5 questions in the entire test booklet, including the core and

---

<sup>4</sup>The bottom 10 percent of respondents were identified by using unweighted data. Because of the nature of the National Assessment of Adult Literacy field-test sample, which combined respondents selected through a probability sample with a convenience sample of respondents selected through focus group centers, the field-test data cannot be weighted to accurately reflect population estimates. However, the field-test data were weighted to roughly reflect the racial/ethnic breakdown of the population (Black, Hispanic, other) and the age breakdown of the population (60 and over vs. under 60). After weighting by these characteristics (race/ethnicity and age), the bottom 10 percent of the field-test sample corresponded to approximately the bottom 5 to 7 percent of the population.

noncore blocks (the fraction comes from partial-credit items) and the 36th respondent correctly answered five questions in the entire test booklet. Thus, to have approximately 10 percent of respondents participate in ALSA, the goal was to identify items that would accurately predict which respondents would correctly answer 5 or fewer of the questions in the core plus the three additional blocks in books 2, 6, or 10.

The following algorithms were developed to discriminate between ALSA and main NAAL respondents by identifying those who would correctly answer five or fewer questions in books 2, 6, or 10:

Algorithm for selecting ALSA respondents for respondents who took the core in English:

- score of 0 on the original 6 core items; OR
- score of 0 on the additional core question (question 9 in block 1 of field-test booklet 2—the flu shot) AND did not attempt to answer core questions 3 and 4.

Algorithm for selecting ALSA respondents for respondents who took the core in Spanish:

- score of less than 5 on the Spanish version of the core; AND
- score of 0 on the additional core question (question 9 in block 1 of field-test booklet 2—the flu shot).

Table 4-7 shows the number of respondents who would be correctly and incorrectly classified into the ALSA and the main NAAL on the basis of this algorithm and the field-test data. Correct classification is defined as placement into the ALSA if the total number of items answered correctly in the entire book (the core plus the three additional blocks) was 5 or fewer and placement into the main NAAL if the total number of items answered correctly in the entire book (the core plus the three additional blocks) was more than 5.

**Table 4-7. Number of field-test respondents correctly and incorrectly classified into the main NAAL and the Adult Literacy Supplemental Assessment (ALSA) on the basis of the algorithm for selecting ALSA respondents: 2003**

Responding category	NAAL	ALSA
Correctly classified respondents	311	32
Incorrectly classified respondents	4	3
Total respondents	315	35

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

On the basis of the recommended criteria, the four respondents who would have been incorrectly screened into the NAAL had total scores of 2, 2.5, 4, and 4 for the cognitive items in the three blocks. The three respondents (0.9 percent of the total sample) misclassified into the ALSA had total scores of 5.5, 5.5, and 13.5. The effect of classifying the two respondents with scores of 5.5 into the ALSA was small because their scores were very close to the cut-point score. According to these criteria for screening respondents into the ALSA, only one of the field-test respondents (the person with a score of 13.5 on the entire booklet) would clearly be misclassified.

### **4.2.3 Analysis of Fluency Addition to the NAAL Field-Test Data**

The Fluency Addition to the NAAL (FAN) field test is described in chapter 5. This section describes the analysis of the FAN field-test data.

Sixteen connected text passages were field-tested for the FAN (eight easy and eight difficult passages) with the goal of selecting eight passages for the operational assessment (four at the easier level and four at the more difficult level). The easier passages included both narratives and expository texts; the more difficult passages were all expository texts.

Table 4-8 shows the results from the FAN field test. Although 16 passages were field-tested, two passages were eliminated because copyright permission to use the passages in the main assessment could not be obtained. A third passage was eliminated because field-test respondents had a lot of difficulty responding to the comprehension question associated with the passage, indicating that the passage was more difficult than intended. Passages in table 4-8 are organized by respondents' average words per minute; easier texts are listed first, followed by more difficult texts. All selected texts were classified as either grade 3 or grade 8 on the basis of Lexile. The column labeled "status" indicates whether the text was kept for the operational assessment or dropped after the field test. The column labeled "explanation" gives the reason for the decision to keep a text. The decision with regard to which passages to include in the operational assessment was based on the following criteria:

- diversity among the passages in terms of structure (narrative vs. expository) and
- diversity among the passages in terms of reading speed.

**Table 4-8. Fluency Addition to the NAAL (FAN) passages, by field-test results: 2003**

Title	Structure	Lexile	Words per minute	Status	Explanation
Curly	Narrative	540 (gr. 3)	186.4	Keep	Fastest narrative, grade 3
Walter	Narrative	590 (gr. 3)	178.1	Drop	Dropped
Rainbow	Narrative	640 (gr. 3)	174.9	Drop	Dropped
Guide Dogs	Expository	700 (gr. 3)	169.4	Keep	Fastest expository text, grade 3
Solar Eclipse	Expository	750 (gr. 4)	167.1	Drop	Dropped
Grand Canyon	Expository	570 (gr. 3)	159.5	Keep	Slow expository text, grade 3
Physical Activity	Expository	650 (gr. 3)	162.5	Drop	Very similar to exercise at grade 8; do not want to spiral two texts with similar topics
Amanda and I	Narrative	700 (gr. 3)	162.1	Keep	Slowest narrative, grade 3
Exercise	Expository (health)	1020 (gr. 8)	164.4	Keep	Fastest passage, grade 8
Lori Goldberg	Expository	1030 (gr. 8)	159.5	Keep	Second fastest passage, grade 8
Elk	Expository	1020 (gr. 8)	147.8	Drop	Dropped
Bigfoot	Expository	1020 (gr. 8)	143.5	Keep	Second slowest passage, grade 8
Chicken Soup	Expository (health)	1100 (gr. 8/9)	136.6	Keep	Slowest passage, grade 8

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

*This page is intentionally left blank.*

## CHAPTER 5

### FIELD TEST OF THE FLUENCY ADDITION TO NAAL

Michelle Amsbary, Westat

#### 5.1 INTRODUCTION

In November 2001, a panel of experts recommended that the government provide, for the first time, a clearer picture of the basic reading skills of low-performing adults by examining their oral reading fluency. In response to this recommendation, an oral reading component for the National Assessment of Adult Literacy (NAAL): the Fluency Addition to the NAAL (FAN) was designed. The FAN field test was conducted to test this newly created oral module assessment and the associated procedures.

In preparation for the field test, the FAN software was developed and tested. The interviewing and data delivery systems and procedures developed for the 2001 NAAL field test were revised to incorporate the newly developed oral module instrument. Instructional manuals and training programs for supervisors and interviewers were developed.

The FAN field test was administered exclusively to volunteers recruited and screened by a focus group facility. The sample design is described in section 5.2. The interviews were conducted by trained interviewers in the respondents' homes.

Interviewers attended a 2-day in-person training that focused on the administration of the oral module but also covered the administration of the screener, the background questionnaire, and the assessment, as well as general contact and administrative procedures. During data collection, two regional supervisors, a field manager, and additional data collection contractor home office staff frequently monitored production. Section 5.3 discusses the data collection instruments and materials, field staff recruitment and training, and the data collection effort for the FAN field test.

Interviewing for the FAN field test lasted 5 weeks, beginning in late September 2002. Interviews were administered in 10 locations across the country by 21 skilled interviewers. A total of 520 interviews were completed.

As described in section 5.4, the data preparation and processing systems and procedures used for the 2001 NAAL field test were refined for the FAN field test. During data collection, data were reviewed for interviewer comments and "other-specify" responses and then converted into the study database. After

frequency review, the data were loaded into the Blaise Editing System, batched, and tracked through delivery to the analysis contractor. The hard-copy assessment booklets were tracked through the Data Management System and sent to the analysis contractor as well. Electronic scoring data from the FAN voice recording files were received from the scoring contractor, reviewed, and delivered for analysis.

## **5.2 SAMPLE DESIGN**

For this purposive field test sample, focus group facilities in the 10 selected locations were contracted to recruit 500 respondents according to specific sampling requirements. These sampling requirements were based on educational attainment, race/ethnicity, and native/nonnative English-speaking ability, as well as a balance of age and gender.

In selecting the sample, sites with high percentages of the target minority populations were chosen, particularly sites where high concentrations of Asian respondents were expected. Minority (especially Asian) populations were targeted to test the FAN instrument with diverse dialects and accents. The focus group services made extraordinary efforts to recruit Asian respondents, including advertising for participants through radio spots and in-person recruiting in public locations, such as shopping malls. Despite these efforts, the Asian respondents were the most difficult to recruit. Toward the end of the field period, the sample requirements were relaxed, allowing the recruitment of additional Asian respondents with a high school diploma in place of those without a high school diploma.

## **5.3 DATA COLLECTION**

### **5.3.1 Data Collection Instruments**

The FAN field test used the following computer-assisted personal interviewing (CAPI) instruments: an abbreviated screener, the background questionnaire, an interviewer guide, and the oral module. These CAPI components were integrated into the Interviewer Management System. A hard-copy assessment booklet containing the core and main assessments was included in the field test as well. These instruments are described below. All instruments were available in both English and Spanish, except for the main exercise, which was available only in English.

The oral module consisted of numerous passages and lists of numbers, letters, and words to be read aloud by the respondent and recorded directly onto the laptop computer. (See section 2.5 for a full discussion of the development and administration of the oral module assessment.) Prior to the field test, a

small feasibility test was conducted with volunteers simply to test the stand-alone oral module. Disks of the recordings created as part of this feasibility test were reviewed.

The version of the background questionnaire used for the FAN field test was the same as that used in the 2001 NAAL field test. Unlike the field test, which used 12 versions of the assessment booklet, the FAN field test used only assessment booklet 2. Because the primary focus of the FAN field test was the oral module, and the items in the assessment booklets had been tested through the 2001 field test, only one type of booklet was used. Additionally, some modifications and enhancements were made to the overall Interviewer Management System to incorporate the new fluency task.

Changes to the Supervisor Management System were implemented to include details about the oral module task. Enhancements were made to the interviewer shipping module (to accommodate the shipping of zip disks), the receipt control system, the browse case functionality, and all study management reports. In addition, the interviewer data transmission process was enhanced to permit the transmission of the FAN administration file. This file collected timing and comment data captured by the oral module.

### **5.3.2 Interviewer Materials**

The materials used in the FAN field test were a subset of those used in the 2001 NAAL field test. The respondents were recruited and screened through a focus group facility, so no advance letter or brochure was required. Additionally, noninterview report forms, segment folders, U.S. Department of Education letters of introduction, community authorization letters, nonresponse letters, and sorry-I-missed-you cards were not used. Handcards were used only for the background questionnaire, not the screener.

The oral module component required numerous materials, including the oral module booklet, the oral module interviewer guide, an interviewer headset, a respondent headset with a microphone, and a zip drive. Four versions of the oral module were tested during the FAN field test. Each version contained the same basic tasks, but the content of the items was slightly different or was presented in a different order. The interviewers' headsets enabled them to hear the signal that the time allotted had expired. (See section 5.3.4.4 for a more detailed discussion of the oral module interviewer guide and interviewer headsets.) The respondent headset with microphone allowed samples of the respondent's reading to be recorded and stored directly on the interviewer laptop. The zip drive was used to create zip disks of the voice recording files. Additionally, procedures were developed for a paper-based backup to the automated scoring

produced from the oral module, in case of problems with the recordings. As part of this procedure, the interviewer used an oral module interviewer guide.

### **5.3.3 Field Staff Recruitment and Training**

Field staff were recruited from among supervisors and interviewers who had experience with the NAAL instrument or a similar literacy instrument. A 2-day field supervisor training was held, followed by a 2-day interviewer training. These training sessions are detailed in section 5.3.3.2.

#### **5.3.3.1 Field Staff Recruitment**

Two experienced supervisors were hired, including one who had supervised interviewers on the 2001 NAAL field test. Twenty-one interviewers were recruited, including several bilingual (Spanish and English) speakers. All interviewers had been employed on either the 2001 field test or a similar literacy-related study.

#### **5.3.3.2 Field Staff Training**

A 2-day supervisor training session was held in mid-September 2002, followed immediately by a 2-day interviewer training session.

New training materials were created for the oral module component. The training materials created for the 2001 NAAL field test were revised to reflect changes resulting from the integration of the oral module, such as the use of the oral module CAPI application and associated booklets, use of the headsets and microphones, creation of zip disk recordings using the zip drive, and the revised contact protocol. A list of procedural and instrumentation changes between the 2001 field test and the FAN field test was developed and distributed to the field staff as well.

The two supervisors were trained on the Supervisor Management System, administrative procedures, use of field production reports, and the CAPI instrumentation. Twenty-one interviewers received extensive training on the administration of the screener, background questionnaire, core and main exercise, and new oral module component.

In addition to interactive and role-play exercises conducted with fellow trainees, interviewers had the opportunity to conduct an interview with a respondent recruited by a local focus group service and paid for his or her participation. This gave the interviewers exposure to a real interviewing environment

before they conducted their first actual case assignment. Software was implemented to permit a quality control review of the audiorecordings captured during the live respondent interviews. All interviews were reviewed and deemed acceptable and usable for analysis, although they were not included in the field test analysis.

#### **5.3.4 Data Collection Effort**

The FAN field test data collection was conducted over 5 weeks in September and October 2002. The field staff structure, progress monitoring, contact protocol, and data collection procedures are presented in the following sections.

##### **5.3.4.1 Field Staff and Organization**

Twenty of the 21 interviewers trained completed the FAN field test. The two supervisors guided and supported the field interviewers working in their region. The supervisors worked closely with interviewers to assign volunteer respondents and monitored progress in scheduling appointments and completing interviews.

The supervisors also assisted the home office study manager in monitoring the focus group service samples and in identifying volunteers requiring replacement. The field director oversaw the work of the supervisors.

##### **5.3.4.2 Monitoring of Data Collection Progress**

Throughout the field period, the FAN sample yield was closely monitored in terms of the number of volunteers recruited and the number of completed background questionnaires. The actual number of completed interviews was compared weekly with the sample requirements for each focus group site.

##### **5.3.4.3 Contact Protocol**

Interviews were conducted in respondents' homes to test the feasibility of administering the FAN in that setting. Interviewers were responsible for contacting the respondent, setting an appointment, getting directions to and locating the respondent's house, making a reminder call, and conducting the interview. Interviews were scheduled at least 3 hours apart, and interviewers were instructed to plan to complete at least two interviews a day.

The focus group facility provided the supervisors with information about the recruited respondents. The supervisors then gave the interviewers a spreadsheet containing information for the respondents in their area, including name, telephone number, address, language preference (English or Spanish), and the best time and day to call.

Using the script on the household folder, the interviewers introduced themselves, set an appointment date and time, and obtained directions to the volunteer's home. Before the appointment, the interviewer made a confirmation call as well. All call attempts were recorded on the Record of Actions on the back of the household folder, as well as in the Electronic Record of Calls in the Interviewer Management System.

#### **5.3.4.4 Data Collection Procedures**

The instruments were administered in the following order: screener, background questionnaire, core and main exercises, and oral module. All instruments were administered in the same visit. Many of the procedures and study materials were similar to or unchanged from those used during the 2001 NAAL field test. Interviewers were not permitted to use bilingual neighbors or other household members to translate, administer, or complete any instruments used in the FAN field test.

In the oral module component, the CAPI system told the interviewer which version of the oral module to administer. The respondent simply read from the oral module booklet, as instructed by the interviewer. During the administration of the oral module, the respondent wore a headset with a microphone so that samples of his or her reading could be recorded. The interviewer wore a headset and followed along in a copy of the booklet used by the respondent, the oral module interviewer guide. The oral module interviewer guide was used as a backup to document the respondent's progress in case the recording equipment failed. Each task was timed, and when the time expired (as indicated by a beep in the interviewer's headset), the interviewer indicated the last word read by the respondent by circling it in the oral module interviewer guide.

#### **5.3.5 Quality Control Measures and Feedback to Staff**

The field manager and supervisors, in addition to the home office staff, monitored production daily. The home office staff also assessed the quality of the data received from the interviewers and provided feedback as necessary. The study hotline staff supported the FAN field test by responding to interviewer calls.

### **5.3.6 Summary of Data Collection Results**

A total of 520 interviews were completed during the FAN field test. All but 13 interviews had an accompanying zip disk containing the respondent voice recordings; all interviews had a paper backup (oral module interviewer guide) available, as needed, for analysis.

The sample requirements were based on race/ethnicity, educational attainment, and the respondent's status as a native or nonnative English speaker.

For race/ethnicity, the goal was to complete interviews with approximately 125 respondents from each of the following groups: Black, Hispanic, Asian, and White. This goal was exceeded in all but the Asian category, which was slightly below the goal.

In the educational attainment category, the goal was to interview 250 respondents in each of two attainment groups (high and low). High educational attainment was defined as having a high school diploma (or general education diploma [GED]) or higher. Low educational attainment was defined as not having a high school diploma or GED. Achieving this goal was complicated by the difficulty in recruiting Asian respondents with less than a high school education until the standard was relaxed to permit recruiting additional Asian respondents with a high school education.

In the native/nonnative English speaker category, the goal was to interview approximately one-fourth of the Hispanic and Asian respondents in each of the following categories: native English speaker with high educational attainment, native English speaker with low educational attainment, nonnative English speaker with high educational attainment, and nonnative English speaker with low educational attainment. The field test came very close to meeting the goal for Hispanic respondents but fell short among Asian respondents, particularly in the case of native English speakers with low educational attainment. Table 5-1 provides further details on completed interviews.

**Table 5-1. Percent of interviews completed in the FAN field test, by race/ethnicity, language status, and educational attainment: 2002**

Language status	Race/ethnicity and educational attainment <sup>1</sup>											
	Black			Hispanic		Asian		White			Other	
	Total	High	Low	High	Low	High	Low	High	Low	Not specified	High	Low
Total	10.0	13.6	11.0	14.2	12.1	14.6	7.3	13.5	11.3	0.2	0.8	1.3
Native English speaker	69.2	13.6	11.0	6.9	5.8	4.4	1.5	13.5	11.0	0.2	0.4	1.0
Nonnative English speaker	30.8	0	0	7.3	6.3	10.2	5.8	0	0.4	0	0.4	0.4

<sup>1</sup> High educational attainment was defined as having a high school diploma (or general education diploma [GED]) or higher. Low educational attainment was defined as not having a high school diploma or GED.

NOTE: All adults of Hispanic origin are classified as Hispanic, regardless of race. Those classified as Black are non-Hispanic Black only. Those classified as Asian are non-Hispanic Asian only. Those classified as White are non-Hispanic White only. Those classified as other include non-Hispanics of all other races, including multiracial.

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

## 5.4 DATA PREPARATION AND PROCESSING

The data preparation and processing activities and procedures implemented for the FAN CAPI instruments and hard-copy materials were very similar to those used for the 2001 NAAL field test. Small revisions were made to the receipt control and editing systems to accommodate the oral module instrument and associated materials. New systems and procedures were developed to support the processing of the zip disks containing the oral module recordings.

### 5.4.1 Receipt, Batching, and Editing of Cases

As cases were transmitted and processed, they were batched into groups of 10. Reports created for each batch provided information on the status of each case: dirty (hard edit encountered), suspect (soft edit encountered), or clean. The reports also included all other-specify entries and remarks entered by interviewers. These reports and the associated cases were reviewed, and data changes were made as appropriate.

The final editing model used for the 2001 NAAL field test, with one additional edit, was used to edit the background questionnaire for the FAN field test. The FAN background questionnaire editing process was primarily focused on reviewing the open-ended responses and interviewer comments.

Edited background questionnaire data were delivered in November 2002, along with the exercise and oral module status codes and the oral module timing data.

#### **5.4.2 Processing and Delivery of Hard-Copy Assessment Materials**

Completed assessment booklets were reviewed by the data collection contractor. They were then shipped to the analysis contractor twice a week throughout the data collection period.

#### **5.4.3 Processing and Delivery of Oral Module Data**

The oral module recordings were copied to zip disks twice a week by the interviewers and mailed to the data collection contractor's home office. Zip disks were receipted and backup copies of the data were made. Completed zip disks were shipped for scoring once a week.

Weekly verification reports were run to ensure that completed oral module files were received for all cases for which the interviewer transmitted a finalized case status. Staff reviewed the scoring contractor's website twice a week to resolve issues with oral module files that could not be processed.

### **5.5 REFINEMENTS FOR MAIN HOUSEHOLD STUDY**

The FAN field test achieved many of its ambitious goals, particularly the development of an instrument to measure basic reading skills and the successful analysis of the related data.

An interviewer debriefing was conducted at the end of the field period to assess interviewer reaction to the training session, strategies to gain respondent cooperation, setup and administration of the oral module instrument, usability of the oral module materials and equipment, and respondent perception of the oral module tasks.

Interviewer input on the wording of the oral module interviewer instructions was obtained as well, in an effort to streamline the instructions, allow more flexibility for the interviewer, and reduce the overall oral module administration time.

These results were used to redesign the FAN, including adding functionality to skip the comprehension questions when the respondent was unable to read the passages and revising the interviewer instructions to make more of the text optional, to be read only when required by the respondent. NCES also identified steps to reduce the oral module administration time, including reducing the number of passages read by the respondent and the number of follow-up questions asked by the interviewer.

*This page is intentionally left blank.*

## CHAPTER 6

### FIELD TEST OF ADULT LITERACY SUPPLEMENTAL ASSESSMENT

Michelle Amsbary and Barbara Forsyth, Westat

The Adult Literacy Supplemental Assessment (ALSA) of the National Assessment of Adult Literacy (NAAL) was developed to gather as much information as possible about adults with limited literacy skills. The ALSA field test was designed to finalize both the ALSA questionnaire and scoring rubrics before their implementation in the NAAL main study. Its principal objective was to ensure that interviewers could score ALSA responses consistently. A three-phase field test for the ALSA was developed. The interviews took place over 3 months, starting in early September 2002. Phase 1 consisted of 10 ALSA interviews, all conducted in English. Phase 2 consisted of 36 ALSA interviews, roughly half conducted in English and the remainder in Spanish. Phase 3 consisted of four interviews, three of which were in Spanish. Six interviewers were recruited and trained for the ALSA field test, including two bilingual interviewers. Each phase consisted of recruiting respondents, training interviewers, and collecting data. Interviewer debriefing sessions were held after each phase.

#### 6.1 PRETEST PHASE 1

The goal of the phase 1 test was to test the English version of the ALSA questionnaire and scoring rubrics developed through cognitive laboratory work. Two interviewers were trained to conduct the phase 1 interviews.

The sample for phase 1 comprised respondents with less than a fifth-grade education in the person's country of origin. Ten volunteers were recruited through adult basic education and literacy programs and literacy tutors in selected areas of Maryland, Virginia, and Washington, D.C. All interviews were conducted in English.

The ALSA stimulus materials included generic materials such as a Coke can or box of pancake mix, as well as materials such as grocery store ads and TV guides that are, by definition, local to only certain geographic areas. These types of materials show some similarities across locales, but also some differences. One of the goals of this pretest was to test whether the similarities were sufficient so that a single material could be used in multiple locales. If the similarities were not sufficient, these types of materials would have to be eliminated from the ALSA assessment.

In response to concerns that familiarity with the stimulus materials might bias the result, “local” and “nonlocal” versions of the questionnaire and stimulus materials were selected and tested for the field test. The local stimulus materials included three items that were specific to the geographic region in which the field test was conducted. The nonlocal stimulus materials included comparable items that were taken from some other geographic region. Each respondent saw only one of the two sets of stimulus materials. Five respondents completed the local version, and four received the nonlocal version. The pilot study was designed to provide evidence of differential performance relative to testing administered with the local versus nonlocal stimulus materials.

The seven core assessment items developed for the NAAL main study were also administered to the ALSA respondents. This approach allowed the survey designers to determine which respondents were true ALSA sample persons based on their performance on the core items.

Interviewers audiotaped all interviews for which respondents gave consent for taping. Observations of as many interviews as feasible were conducted. Written summaries of the observed interviews were produced. Both the audiotapes and written summaries were used to clarify procedures for subsequent testing.

Interviewers completed nine interviews, using the English language version of the questionnaire. The phase 1 results were used to refine the questionnaire and scoring rubrics and to revise training materials for field-test phases 2 and 3.

On the basis of performance on the core assessment, five of the nine respondents were determined to have been appropriate ALSA candidates; the remainder would not have been eligible for the ALSA.

## **6.2 PRETEST PHASE 2**

Phase 2 was the first test with revised Spanish language materials after the cognitive laboratory work. It was anticipated that changes to the Spanish language materials and interview procedures would be greater than changes to the English language materials and procedures. Six interviewers were trained to conduct this pretest, two of whom were bilingual.

The data collection contractor visited adult basic education programs in Montgomery County, Maryland, to recruit adults known by teachers to have relatively low reading test scores appropriate for the ALSA assessment. Work establishments were also visited to recruit Spanish-speaking volunteers with less than a fifth-grade education. These visits were followed up with telephone calls to schedule interview

appointments. Fifty adults were recruited; of these, 40 interviews were scheduled, including 20 with English-only speakers and 20 with Spanish-speaking volunteers.

On the basis of the results of the phase 1 test, the ALSA questionnaire and interviewing methods were revised slightly for phase 2. The changes were incorporated into both the English and Spanish versions of the instruments.

Thirty-six interviews were conducted, including 17 in Spanish. All interviews were successfully tape recorded. Sixteen interviews were observed.

The average administration length for the ALSA assessment in phase 2 was 50.9 minutes. The core assessment was administered in 34 of the 36 completed interviews. Of these 34 respondents, 15 passed the core assessment and 19 failed. Therefore, these 19 were true ALSA respondents and would have been classified into the ALSA according to the criteria used for the NAAL main study.

The phase 2 results were used to revise interview procedures and to further refine the scoring rubrics, interviewer materials, and training materials.

### **6.3 PRETEST PHASE 3**

The goal of the small phase 3 study was to test any final revisions to interview materials and procedures before finalizing them for the NAAL main study. One bilingual interviewer with experience in the previous phases of the pretests conducted these final interviews.

In preparation for phase 3, the ALSA referral sources were reviewed to determine which were most effective in providing truly eligible ALSA respondents. It was determined that in phases 1 and 2, literacy programs provided the highest respondent eligibility rates (71 percent), followed by sources such as internal recruiting efforts and nearby business establishments (69 percent), and Montgomery County, Maryland, adult basic education programs (36 percent). This information was used to guide the recruitment effort for phase 3. Five respondents were recruited for this phase.

On the basis of phase 2 field test observations and the interviewer debriefing, several refinements were made to the ALSA questionnaire. The largest questionnaire revision involved adding scripted probes to the vocabulary items to assist interviewers when respondents gave no response. Minor revisions were made to clarify some of the rubrics and the Spanish language translations.

Four interviews (three in Spanish and one in English) were conducted during this final phase.

## **6.4 RESULTS OF THE FIELD TEST**

A total of 49 interviews were completed during the three phases of the ALSA data collection. The data collection contractor observed the interviews and submitted written summaries of the observations, along with copies of the completed questionnaires and interview audiotapes, for review and analysis.

Results from all three phases were used to determine whether nonlocal materials could be effective in the main data collection effort. On the basis of the three phases of interviewing, it was decided to use nonlocal stimulus materials in the main data collection effort. The questionnaire content was revised to accommodate the selected stimulus materials, and preparations were made for the NAAL main study.

Additionally, the timing data indicated that the ALSA instrument needed to be shortened so that respondents would not be forced to struggle for an extended time, possibly becoming frustrated before completing the Fluency Addition to NAAL. A recommendation was made to shorten the ALSA instrument by eliminating six vocabulary items, five items asking respondents to read connected text, and the telephone book stimulus.

## CHAPTER 7

### SAMPLE DESIGN

Leyla Mohadjer and Thomas Krenzke, Westat

#### 7.1 INTRODUCTION

The purpose of the National Assessment of Adult Literacy (NAAL) was to estimate the literacy levels of the adult population in the United States. To adequately represent the target population, the NAAL included both a household component and a prison component. The NAAL household study included two sets of household samples: a national household sample and household samples from six states. The target population for the national and state household samples consisted of adults 16 years or older who resided in housing units at the time of interview. This population of 221,020,000 adults in 2003 included persons who resided in college dormitories but excluded adults in military barracks, halfway houses, and other group quarters.<sup>1</sup> Each household sample was selected on the basis of a four-stage, stratified cluster sample.

As in the 1992 National Adult Literacy Survey (NALS), all states were given an opportunity to explore the skill levels of their adults by participating in the State Assessment of Adult Literacy (SAAL) part of the NAAL. (Participating states paid the cost of the additional assessments.) The states participating in SAAL were Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma. The NAAL and SAAL samples were combined (composited) to improve the precision of statistics for the national and SAAL samples. A prison component was added to improve the representation of the target population. The prison component of NAAL included a sample of adult inmates in federal and state prisons, which is representative of the 1,380,000 adults in prisons in 2003. Together, the household and prison samples are representative of the 222,400,000 adults in American households and prisons.

Section 7.2 provides a detailed summary of the four stages of sampling for the household samples. Section 7.3 describes the two-stage sample design for the prison sample. Section 7.4 discusses the assignment of booklet types within the household and prison samples.

---

<sup>1</sup> All people not living in housing units are classified as living in group quarters. There are two general categories: (1) institutionalized population (such as nursing homes and schools, hospitals, and wards for the mentally retarded) and (2) noninstitutionalized population (such as religious group quarters or emergency and transitional shelters).

## **7.2 HOUSEHOLD SAMPLES**

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

Section 7.2.1 presents the key features of the household samples and summarizes sample sizes at each sampling stage. Section 7.2.2 discusses the target population and the information sources used to create the frames for each stage of sampling. The selection process for each of the four stages is described in section 7.2.3.

### **7.2.1 Key Design Features and Sizes of the Household Samples**

A single area sample was selected for the national NAAL sample, and six additional SAAL samples were selected for Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma. For each sample, the usual procedures for area sampling were followed: a stratified probability proportionate to size design was used for the first two stages and systematic random samples were drawn in the last two stages.

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

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

The first stage of sampling was the selection of PSUs, which consisted of counties or groups of counties. PSUs were formed within state boundaries, which gave an improved sample for state-level estimation. One PSU was selected per stratum by using probabilities proportionate to their population within households, except in Maryland and Massachusetts where samples of segments were selected as

the first-stage units (refer to section 7.2.3.1.1 for more discussion). One hundred PSUs were selected for the national sample, and 54 PSUs were selected in Kentucky, Missouri, New York, and Oklahoma. In Maryland and Massachusetts there were too few PSUs to sample from; therefore segments were selected in the first stage of sampling. After selecting the segments, 20 area clusters (quasi-PSUs) were created for Maryland and Massachusetts by grouping the selected segments into 20 geographically clustered areas to facilitate a cost-efficient approach to data collection. The true first-stage sample size is much larger because a total of 323 first-stage units (i.e., segments) were selected in Maryland and Massachusetts. However, to not mislead readers into thinking the sample is much larger than it really is, this report uses the PSU count of 20 in Maryland and Massachusetts to describe the first-stage sampling activities. Fourteen PSUs were selected for both the national NAAL and the SAAL samples; hence, the sample included a combined total of 160 unique PSUs.

In the second stage of sampling, segments (census blocks or groups of blocks) within the PSUs were selected with a probability proportionate to size; the measure of size (MOS) for a segment was a function of the number of year-round housing units within the segment. In the national sample, the Black and Hispanic populations were sampled at a higher rate than the remainder of the population to increase their sample size, whereas the state samples used no oversampling. Oversampling in the national sample was accomplished by oversampling the high-minority segments in which Black and Hispanic adults accounted for 25 percent or more of the population. There were 1,959 segments selected for the national sample and 861 segments selected across the SAAL samples, with a total of 2,818 unique segments selected across the national and six SAAL samples. (Two segments were selected for both the NAAL and SAAL samples.)

In the third stage of sampling, housing units were selected with equal probability within each segment, except for nonminority households within high-minority segments in the national component. These national sample households were subsampled after screening so that the sampling rates for nonminority persons would be about the same in the high-minority segments as in other segments. The overall sample size of housing units took into account expected losses owing to vacant housing units, units that were not housing units, and expected response rates.<sup>2</sup>

---

<sup>2</sup> The expected response rates took into account those experienced in the 1992 NALS and recent trends in household studies.

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

Table 7-1 contains the sample sizes of PSUs, segments, housing units, and persons and the number of persons completing the background questionnaire for the national NAAL household sample.

**Table 7-1. NAAL sample sizes of PSUs, segments, housing units, and persons and numbers of completed background questionnaires, by region, Metropolitan Statistical Area (MSA) status, and segment status: 2003**

Characteristic	Primary sampling units (percent)	Segments	Housing units	Persons	Completed background questionnaires <sup>1</sup>
Total	100	1,959	24,450	16,409	12,753
Region <sup>2</sup>					
Northeast	18	373	4,510	2,864	2,228
Midwest	23	404	4,833	3,354	2,688
South	38	758	10,295	6,405	4,943
West	21	424	5,812	3,786	2,894
MSA status					
Non-MSA	22	357	4,282	2,862	2,295
MSA	78	1,602	21,168	13,547	10,458
Segment status <sup>3</sup>					
Low minority	†	1,091	11,648	8,441	6,394
High minority	†	868	13,802	7,968	6,359

† Not applicable.

<sup>1</sup> Completed background questionnaires included cases that were not complete due to language problems and mental disabilities. These cases were considered a “success” in data collection since race/ethnicity, age, and gender were collected, as well as good information (language problem or mental disability) as to their English literacy skills.

<sup>2</sup> Northeast Region = ME, NH, VT, MA, RI, CT, NY, NJ, PA; Midwest Region = OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS; South Region = DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX; West Region = MT, ID, WY, CO, NM, AZ, UT, NV, WA, OR, CA, AK, HA.

<sup>3</sup> Segments classified as low minority have less than 25 percent Black and Hispanic adults. Segments classified as high minority have 25 percent or more Black and Hispanic adults. The cutpoint of 25 percent of the population being Black and Hispanic was derived analytically for the 1992 NALS. An analysis for the 1992 sample design showed that the 25 percent concentration provided the minority sample sizes at acceptable design effect levels. The same cutpoint was used for the 2003 NAAL study.

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

The SAAL design called for a sample of 12 PSUs in each state. The first stage of sampling for Maryland and Massachusetts was the selection of segments because these states had small numbers of PSUs (i.e., Maryland and Massachusetts had three stages of sampling instead of four; refer to section 7.2.3.1.1 for a discussion of the selection of segments as the first-stage units in Maryland and Massachusetts). Eighteen PSUs instead of 12 were selected for Kentucky to improve the precision of small-area estimation as requested by this state. Twelve PSUs were selected for Missouri, New York, and Oklahoma.

The NAAL and SAAL samples were integrated through a compositing procedure to achieve the maximum efficiency for producing both national and state estimates for the participating states. Table 7-2 summarizes the sample sizes for the combined NAAL-SAAL sample, as well as the numbers of completed background questionnaires. More information about sample yields can be found in the tables in section 8.8.3.

**Table 7-2. NAAL-SAAL combined sample sizes of PSUs, segments, housing units, and persons and numbers of completed background questionnaires, by region, Metropolitan Statistical Area (MSA) status, and state: 2003**

Characteristic	Primary sampling units <sup>1</sup>	Segments	Housing units	Persons	Completed background questionnaires <sup>2</sup>
Total	174	2,820	35,365	23,732	18,541
Total unduplicated	160	2,818	35,365	23,732	18,541
Region <sup>3</sup>					
Northeast	37	636	7,518	4,936	3,765
Midwest	35	538	6,332	4,500	3,612
South	81	1,222	15,703	10,510	8,270
West	21	424	5,812	3,786	2,894
MSA status					
Non-MSA	59	584	6,948	4,835	3,897
MSA	115	2,236	28,417	18,897	14,644
State					
Kentucky	20	229	2,696	1,945	1,545
Maryland	14	150	1,727	1,276	1,016
Massachusetts	9	173	2,002	1,432	1,074
Missouri	13	149	1,635	1,247	1,009
New York	20	310	3,680	2,251	1,730
Oklahoma	14	164	1,924	1,533	1,287
Rest of U.S.	84	1,645	22,701	14,048	10,880

<sup>1</sup> Because segments were the first stage of sampling for Maryland and Massachusetts, PSU counts for these states represent groups of segments formed for field management purposes.

<sup>2</sup> Includes respondents who did not complete the background questionnaire because of language problems or mental disabilities.

<sup>3</sup> Northeast Region = ME, NH, VT, MA, RI, CT, NY, NJ, PA; Midwest Region = OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS; South Region = DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX; West Region = MT, ID, WY, CO, NM, AZ, UT, NV, WA, OR, CA, AK, HA.

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

## 7.2.2 Sampling Frames for the Household Study

The target population for the national and state household studies consisted of adults 16 years of age or older who resided in housing units at the time of interview. Area sampling methodology was used to facilitate the selection of a representative sample from the target population. Area sampling requires the formation of frames at each stage of sampling. The next subsections discuss the creation of sampling frames for each stage of selection.

### 7.2.2.1 Sampling Frame for Primary Sampling Units

For the initial stage of sampling, a PSU frame was created by using the census 2000 public law (PL)-94 county-level files. The PSUs were formed by combining adjacent counties, respecting their population sizes and taking into consideration the travel distance for interviewers. The PSUs were formed as a single county or a group of contiguous counties, depending on the population size and the end-to-end distance within a PSU.

One set of PSUs was created for both the NAAL and the SAAL household samples. The objective of the PSU formation process was to minimize travel distance within a PSU (where the maximum distance was 100 miles), subject to a minimum population size (i.e., a minimum MOS) in a PSU of 15,000. The census 2000 PL-94 county-level data were used to obtain county-level population sizes. The MOS variable was an estimate of the population counts within households derived from the total population counts from the 2000 census.

The PSUs were formed within MSA boundaries. They also were formed within state boundaries, with the exception of two PSUs that were not selected in either the NAAL sample or the SAAL sample. A total of 1,884 PSUs were formed. Table 7-3 presents the distribution of counts of PSUs by region and MSA status, and table 7-4 presents some characteristics of the PSUs by MSA status. The PSU frame for SAAL states was created by subsetting the PSUs from the NAAL PSU frame. For states where PSUs were selected as the first stage, the numbers of PSUs formed before sampling were 73 (Kentucky), 65 (Missouri), 39 (New York), and 45 (Oklahoma).

**Table 7-3. Distribution of formed primary sampling units, by Metropolitan Statistical Area (MSA) and region status: 2003**

Region <sup>1</sup>	Total	Non-MSA	MSA
Total	1,884	1,497	387
Northeast	143	88	55
Midwest	610	510	100
South	853	695	158
West	278	204	74

<sup>1</sup> Northeast Region = ME, NH, VT, MA, RI, CT, NY, NJ, PA; Midwest Region = OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS; South Region = DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX; West Region = MT, ID, WY, CO, NM, AZ, UT, NV, WA, OR, CA, AK, HA.

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

**Table 7-4. Distributions of primary sampling unit characteristics, by Metropolitan Statistical Area (MSA) status, mean, median, and percentiles: 2003**

Variable	Category	Mean	Median	First quartile	Third quartile	95th percentile
Population size	Non-MSA	36,344	29,441	20,479	42,608	83,340
	MSA	577,987	255,403	134,309	627,018	2,125,203
	Total	147,605	35,784	22,354	73,997	650,700
Distance (miles) <sup>1</sup>	Non-MSA	60.3	46.8	37.3	65.1	129.8
	MSA	70.6	62.3	47.4	79.8	115.9
	Total	62.4	49.3	38.3	69.3	127.7
Area (square miles)	Non-MSA	1,878	823	555	1,422	6,082
	MSA	1,840	1,364	805	2,097	4,204
	Total	1,870	900	572	1,716	5,660
Number of counties	Non-MSA	1.5	1.0	1.0	2.0	3.0
	MSA	2.2	2.0	1.0	3.0	5.0
	Total	1.7	1.0	1.0	2.0	4.0

<sup>1</sup> The distance was computed as the end-to-end distance.. It is beneficial to minimize this distance in order to minimize travel distance for interviewers.

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

### 7.2.2.2 Sampling Frame for Segments

For the second stage of sampling, a frame of segments was created within the selected PSUs by using the census 2000 summary file (SF) block data. A segment consists of a census block<sup>3</sup> (as defined by Census 2000) or a combination of two or more nearby blocks. Within each PSU, the block data from the 2000 census SF1 files were sorted by tract, block group, and block number before creating the segments.

Blocks with no housing units and no population were also included in the formation process in order to include all housing units constructed after the 2000 census. A single block was used as a segment when the number of housing units in the block exceeded 60. Neighboring blocks were combined within a tract to reach either the required minimum of 60 housing units per segment or the end of the tract (segments did not cross tract boundaries).

<sup>3</sup> Blocks are very fine partitions of the United States, formed by using visible semipermanent features such as roads, railroad tracks, mountain ridges, bodies of water, and power lines. The only invisible boundaries used are county, state, and national boundaries. Minor civil division boundaries and property lines are ignored. A block group is a small group of contiguous blocks. A tract is a collection of contiguous block groups all within the same county.

A total of 413,523 segments were formed for the NAAL PSUs, including NAAL PSUs that overlapped with the SAAL PSUs. Another 39,769 segments were formed within PSUs selected for the SAAL only. Therefore, a total of 453,292 segments were formed for the combined NAAL-SAAL PSU sample.

### 7.2.2.3 Sampling Frame for Housing Units

After segments were selected, the data collection contractor’s listers visited each sampled segment to create a sampling frame of housing units for the third stage of sampling. Interviewers constructed a list of all housing units within the segment boundaries, using tract and segment maps created by home office staff. A small number of segments were subdivided, with one part, or “chunk,” selected at random for listing. Chunking reduced the burden of listing large sampled segments (generally more than 300 housing units) by dividing the segment into chunks. A chunk was selected with probability proportionate to the estimated number of housing units within the chunk, and listing was conducted within the selected chunk. Of the 2,818 segments selected, 334 were sampled chunks of segments. Table 7-5 shows the distribution of the segments by size.

**Table 7-5. Percentile distributions of NAAL-SAAL sampled segments: 2003**

Percentile	Actual number of housing units listed <sup>1</sup>
5th	59
25th	71
50th	98
75th	154
95th	238
Mean	118

<sup>1</sup> Counts reflect number of housing units listed, whether from a chunk or a full segment.

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

### 7.2.2.4 Sampling List of Persons

At the fourth stage of sampling, households were screened to determine whether they included any eligible persons, defined as a person 16 years or older who resided in the household; the definition of eligibility included persons who resided in college dormitories but excluded adults in military barracks, halfway homes, and group quarters. A complete list of household members was obtained by interviewers

and entered in the CAPI program as part of the screener interview conducted in each sampled household. Before sample selection, the CAPI system determined the eligibility of each person listed.

### **7.2.3 Sample Selection for the Household Study**

The NAAL sample and the SAAL sample were selected independently on the basis of a four-stage, stratified cluster sample involving the selection of PSUs, segments, housing units, and eligible persons within households in the selected housing units. The selection of PSUs is discussed in section 7.2.3.1. The selection of segments within PSUs and of housing units within sampled segments is discussed in sections 7.2.3.2 and 7.2.3.3, respectively. The sampling of persons is discussed in section 7.2.3.4. Section 7.2.3.5 describes an approach to improve the coverage of the housing unit sampling frame, and section 7.2.3.6 discusses the probabilities of selection. Differences between the NAAL and SAAL sampling procedures are pointed out within each section. Selection results are summarized for the NAAL sample, the SAAL sample, and the combined NAAL-SAAL sample.

#### **7.2.3.1 Selection of Primary Sampling Units**

The PSUs were selected as a stratified probability-proportionate-to-MOS sample, where the MOS was equal to the household population. The PSUs were then stratified on the basis of available variables from the census 2000 PL-94 files. The stratification is described in section 7.2.3.1.1, and the selection process is discussed in section 7.2.3.1.2.

##### **7.2.3.1.1 Stratification of Primary Sampling Units**

The NAAL and SAAL samples involved the selection of PSUs, with one PSU selected per stratum. For this selection, the PSUs on the frame were stratified into homogeneous strata. The PSUs with the largest MOS were selected with certainty. Each of the certainty PSUs was treated as a single stratum, and the remaining PSUs were stratified into the appropriate number of noncertainty strata.

The certainty PSUs were identified before the application of the stratification algorithm. The certainty PSUs were the largest PSUs (in terms of MOS) and were selected with probability equal to 1. The certainty cutoff was determined from probability proportionate to size sampling, with the total population in households as the MOS.

On the basis of the analytical cutoff, PSUs were selected with certainty independently for the NAAL and the SAAL samples. The remaining PSUs for each respective sample, excluding the certainty

PSUs, were stratified into noncertainty strata. The NAAL sample stratification process initially formed major strata defined by census division and MSA status (except in the New England and Mid-Atlantic census divisions, where non-MSAs were combined into one major stratum).

Each major stratum was further stratified into the allocated number of substrata. The main objective of the substratification process was to keep the substratum (i.e., ultimate stratum) sizes as equal as possible, both to reduce the variation in workload and to control the variances of the estimates. Table 7-6 presents the variables used for NAAL substratification within each major stratum. The variables used in the substratification process were identified through a regression analysis. The dependent variable for the stepwise regression analysis was the percentage of the population that were high school graduates 25 years and older. Limited by the time that the stratification occurred (just after Census 2000), the independent variables were census division, MSA status, per capita income, percentage of Non-Hispanic black population, percentage of Hispanic population, percentage of non-minority population, and the PSU population size.

The sample designs for Maryland and Massachusetts required one fewer sampling stage than the designs for the other SAAL states and the NAAL sample. Because the numbers of PSUs formed within Maryland and Massachusetts were small, a sample of segments was selected across each state. Within each of these two states, segments were sampled with probability proportionate to size. Table 7-7 presents the variables used for the SAAL substratification process. The variables used in the substratification process were identified through a stepwise regression analysis, using the same variables as described in the preceding paragraph for the NAAL stratification.

**Table 7-6. Variables used in NAAL noncertainty primary sampling unit stratification: 2003**

Census division	Metropolitan Statistical Area (MSA) status	Stratification variables
New England Division (1) and Middle Atlantic Division (2)	Non-MSA	Per capita income
East North Central Division (3) and West North Central Division(4)	Non-MSA	Per capita income, percentage nonminority
South Atlantic Division (5) East South Central Division (6) and West South Central Division (7)	Non-MSA	Per capita income, percentage non-Hispanic Black
Mountain Division (8) and Pacific Division (9)	Non-MSA	Per capita income
New England Division (1) and Middle Atlantic Division (2)	MSA	Per capita income, percentage Hispanic
East North Central Division (3)	MSA	Per capita income, percentage non-Hispanic Black
West North Central Division (4)	MSA	Per capita income
South Atlantic Division (5)	MSA	Per capita income, percentage non-Hispanic Black, percentage Hispanic
East South Central Division (6)	MSA	Per capita income, percentage non-Hispanic Black
West South Central Division (7)	MSA	Per capita income, percentage non-Hispanic Black, percentage Hispanic
Mountain Division (8) and Pacific Division (9)	MSA	Per capita income, percentage nonminority

NOTE: New England Division (1) = ME, NH, VT, MA, RI, and CT; Middle Atlantic Division (2) = NY, NJ, and PA; East North Central Division (3) = OH, IN, IL, MI, and WI; West North Central Division (4) = MN, IA, MO, ND, SD, NE, and KS; South Atlantic Division (5) = DE, MD, DC, VA, WV, NC, SC, GA, and FL; East South Central Division (6) = KY, TN, AL, and MS; West South Central Division (7) = AR, LA, OK, and TX; Mountain Division (8) = MT, ID, WY, CO, NM, AZ, UT, and NV; and Pacific Division (9) = WA, OR, CA, AK, and HI.

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

**Table 7-7. Variables used in SAAL noncertainty primary sampling unit stratification: 2003**

State	Metropolitan Statistical Area (MSA) status	Stratification variables
New York	Non-MSA	Per capita income
	MSA	Per capita income, percent Hispanic
Missouri	Non-MSA	Per capita income, percent nonminority
	MSA	Per capita income
Kentucky and Oklahoma	Non-MSA	Per capita income, percent non-Hispanic Black
	MSA	Per capita income, percent non-Hispanic Black, percentage Hispanic

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

### 7.2.3.1.2 PSU Selection

One PSU was selected independently in each stratum for the NAAL and SAAL samples. A PSU in a certainty stratum had a selection probability of 1, and the PSUs in noncertainty strata were selected with probability proportional to measure of size (MOS) (i.e., the total population within households). The formula for the PSU selection probability, involving the PSU MOS, is provided in table 7-8.

**Table 7-8. NAAL-SAAL selection probabilities: 2003**

Sampling unit PSUs	Conditional probability <sup>†</sup>	Overall probability
	<p>†</p>	<p>Let <math>x_{hi}</math> = total population in households for PSU <math>i</math> within stratum <math>h</math>.</p>
	<p>Let <math>HU_{Nij}</math> = number of nonminority housing units (HUs) expected in segment <math>j</math> of PSU <math>i</math>.</p> <p>Let <math>HU_{Mij}</math> = number of minority housing units expected in segment <math>j</math> of PSU <math>i</math>.</p> <p>Let <math>MOS_{1ij}</math> = <math>HU_{Nij} + 3HU_{Mij}</math>, if high-minority NAAL segment.<sup>1</sup></p> <p>Let <math>MOS_{2ij}</math> = <math>HU_{Nij} + HU_{Mij}</math>, if low-minority NAAL segment.<sup>1</sup></p> <p>Let <math>MOS_{3ij}</math> = <math>HU_{Nij} + HU_{Mij}</math>, if SAAL segment.</p> <p>For high-minority NAAL segments, the conditional probability is</p> $CP_{ij} = \frac{n \left[ \frac{MOS_{1ij}}{P_{hi}} \right]}{\sum_{ij} \left[ \frac{MOS_{1ij}}{P_{hi}} \right] + \sum_{ij} \left[ \frac{MOS_{2ij}}{P_{hi}} \right]} = \frac{MOS_{1ij} / P_{hi}}{I_{NAAL}}$ <p>and for low-minority NAAL segments, the conditional probability is</p> $CP_{ij} = \frac{n \left[ \frac{MOS_{2ij}}{P_{hi}} \right]}{\sum_{ij} \left[ \frac{MOS_{1ij}}{P_{hi}} \right] + \sum_{ij} \left[ \frac{MOS_{2ij}}{P_{hi}} \right]} = \frac{MOS_{2ij} / P_{hi}}{I_{NAAL}}$ <p>where <math>n</math> = total size of the target segment sample for NAAL. For SAAL segments, the conditional probability is</p> $CP_{ij} = \frac{n \left[ \frac{MOS_{3ij}}{P_{hi}} \right]}{\sum_{ij} \left[ \frac{MOS_{3ij}}{P_{hi}} \right]} = \frac{I_{SAAL}}{I_{SAAL}}$ <p>where <math>n</math> = total size of the target segment sample for SAAL.</p>	$P_{hi} = \frac{x_{hi}}{\sum_{i \in h} x_{hi}}$
		<p>For high-minority NAAL segments:</p> $P_{ij} = P_{hi} \times \frac{MOS_{1ij} / P_{hi}}{I_{NAAL}} = \frac{MOS_{1ij}}{I_{NAAL}}$ <p>For low-minority NAAL segments:</p> $P_{ij} = P_{hi} \times \frac{MOS_{2ij} / P_{hi}}{I_{NAAL}} = \frac{MOS_{2ij}}{I_{NAAL}}$ <p>For SAAL segments:</p> $P_{ij} = P_{hi} \times \frac{MOS_{3ij} / P_{hi}}{I_{SAAL}} = \frac{MOS_{3ij}}{I_{SAAL}}$

See notes at end of table.

**Table 7-8. NAAL-SAAL selection probabilities: 2003—Continued**

Sampling unit	Conditional probability	Overall probability
Chunking segments and NAAL/SAAL overlap segments	<p>Let <math>HU_{ij}(chunk)</math> = number of housing units expected in the selected chunk in segment <math>j</math> of PSU <math>i</math>.</p> <p>Let <math>HU_{ij}</math> = number of housing units expected in segment <math>j</math> of PSU <math>i</math>.</p> $CP_{ij}(chunk) = \begin{cases} \frac{HU_{ij}(chunk)}{HU_{ij}}, & \text{if chunking is implemented.} \\ 1, & \text{otherwise.} \end{cases}$	<p>For high-minority NAAL segments:</p> $P_{ij}(chunk) = \frac{MOS_{1ij}}{I_{NAAL}} \times \frac{HU_{ij}(chunk)}{HU_{ij}}.$ <p>For low-minority NAAL segments:</p> $P_{ij}(chunk) = \frac{MOS_{2ij}}{I_{NAAL}} \times \frac{HU_{ij}(chunk)}{HU_{ij}}.$ <p>For SAAL segments:</p> $P_{ij}(chunk) = \frac{MOS_{3ij}}{I_{SAAL}} \times \frac{HU_{ij}(chunk)}{HU_{ij}}.$
Housing units	<p>Let <math>\bar{d}</math> = average number (integer) of housing units to be sampled within NAAL segments. Let <math>\bar{d}_{state}</math> = average number (integer) of housing units to be sampled within SAAL segments. Then the within-segment sampling rate (i.e., conditional selection probability) is</p> $CP_{ijk} = \frac{3\bar{d} / CP_{ij}(chunk)}{HU_{ij}} \text{ for high-minority NAAL segments,}$ $CP_{ijk} = \frac{\bar{d} / CP_{ij}(chunk)}{HU_{ij}} \text{ for low-minority NAAL segments, and}$ $CP_{ijk} = \frac{\bar{d}_{state} / CP_{ij}(chunk)}{HU_{ij}} \text{ for SAAL segments within a state.}$	<p>For housing units in high-minority NAAL segments:</p> $P_{ijk} = \frac{MOS_{1ij}}{I_{NAAL}} \times \frac{HU_{ij}(chunk)}{HU_{ij}} \times \frac{3\bar{d} / CP_{ij}(chunk)}{MOS_{1ij}}$ $= \frac{3\bar{d}}{I_{NAAL}}.$ <p>For housing units in low-minority NAAL segments:</p> $P_{ijk} = \frac{MOS_{2ij}}{I_{NAAL}} \times \frac{HU_{ij}(chunk)}{HU_{ij}} \times \frac{\bar{d} / CP_{ij}(chunk)}{MOS_{2ij}}$ $= \frac{\bar{d}}{I_{NAAL}}.$

See notes at end of table.

**Table 7-8. NAAL-SAAL selection probabilities: 2003—Continued**

Sampling unit	Conditional probability	Overall probability
Housing units—Continued		<p>For SAAL housing units:</p> $P_{ijk} = \frac{MOS_{3ij}}{I_{SAAL}} \times \frac{HU_{ij(chunk)}}{HU_{ij}} \times \frac{\bar{d}_{state} / CP_{ij(chunk)}}{MOS_{3ij}}$ $= \frac{\bar{d}_{state}}{I_{SAAL}}$
Missed housing units	<p>If a large number of housing units were found during either the missed structure or hidden housing unit procedure, a sample of <i>mhu</i> missed housing units was selected from the <i>MHU</i> missed housing units identified. This occurred for the missed structure process only.</p> $CP_{ijk(mhu)} = \begin{cases} \frac{mhu_{ijk}}{MHU_{ijk}} & \text{for missed HU sample or} \\ 1, & \text{otherwise.} \end{cases}$	<p>For NAAL housing units:</p> $P_{ijk(mhu)} = \frac{\bar{d}}{I_{NAAL}} \times \frac{mhu_{ijk}}{MHU_{ijk}}$ <p>For SAAL housing units:</p> $P_{ijk(mhu)} = \frac{\bar{d}_{state}}{I_{SAAL}} \times \frac{mhu_{ijk}}{MHU_{ijk}}$
Persons	<p>Let <i>M</i> = number of eligible persons within a household. If <i>M</i> ≤ 3, then <i>m</i> = 1. If <i>M</i> &gt; 3, then <i>m</i> = 2.</p> $CP_{ijkl} = \frac{m_{ijk}}{M_{ijk}}$	<p>For NAAL:</p> $P_{ijkl} = \frac{\bar{d}}{I_{NAAL}} \times \frac{mhu_{ijk}}{MHU_{ijk}} \times \frac{m_{ijk}}{M_{ijk}} \times 0.95,$ <p>where the factor 0.95 is applied because the reserve sample was not released.</p> <p>For SAAL:</p> $P_{ijkl} = \frac{\bar{d}}{I_{SAAL}} \times \frac{mhu_{ijk}}{MHU_{ijk}} \times \frac{m_{ijk}}{M_{ijk}} \times 0.95.$

† Not applicable.

‡ High-minority segments are segments with 25 percent or more Blacks and Hispanics; otherwise segments are classified as low-minority segments.

NOTE: h = stratum; i = PSU; j = segment; k = HU; l = person; Phi is the overall probability of selection for PSU i within stratum h; P<sub>ij</sub> is the overall probability of selection for segment j within PSU i;

INAAL is the sampling interval for the selection of the NAAL segments; ISAAL is the sampling interval for the selection of the SAAL segments.

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

For the NAAL sample, 100 PSUs were selected. One PSU was selected from each of 100 (16 certainty and 84 non-certainty) strata. Table 7-1 provides the sample sizes of PSUs by census region and MSA status. Table 7-9 shows the number of PSUs selected in each sample by certainty status.

**Table 7-9. Distribution of sampled PSUs, by certainty status: 2003**

Sample	Total <sup>1</sup>	PSU certainty status	
		Noncertainty	Certainty
NAAL	100	84	16
SAAL			
Kentucky	18	15	3
Maryland	13	†	†
Massachusetts	7	†	†
Missouri	12	8	4
New York	12	4	8
Oklahoma	12	9	3

† Not applicable.

<sup>1</sup> Because segments were the first stage of sampling for Maryland and Massachusetts, PSU counts for these states represent groups of segments formed for field management purposes.

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

The NAAL and SAAL PSU samples were selected independently. As a result, 14 PSUs were selected for both the NAAL and the SAAL (many of them being certainty PSUs). Table 7-10 presents the number of overlapping PSUs by SAAL state.

**Table 7-10. Number of overlapping PSUs in the NAAL and SAAL samples: 2003**

State	Number of overlapping PSUs <sup>1</sup>
Total	14
Kentucky	1
Maryland	1
Massachusetts	2
Missouri	1
New York	7
Oklahoma	2

<sup>1</sup> Because segments were the first stage of sampling for Maryland and Massachusetts, PSU counts for these states represent groups of segments formed for field management purposes.

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

### **7.2.3.2 Selection of Segments**

To increase the number of Black and Hispanic adults in the NAAL sample, segments with moderate to high concentrations of Black and Hispanic adults were given a higher selection probability. Segments in which Blacks or Hispanics accounted for 25 percent or more of the population were oversampled at a rate up to three times that of the remainder of the segments. The housing unit counts served as the MOS for the low-minority segments (segments in which Blacks and Hispanics accounted for less than 25 percent of the population). In high-minority segments, the MOS was the number of White, non-Hispanic households plus three times the number of Black and Hispanic households. Table 7-8 shows the MOS and the probabilities of selection for the second-stage sample. A minimum MOS of 5 was assigned to each segment with fewer than five housing units. Ordering the frame of segments within each PSU by the proportion of Blacks and Hispanics in the segment provided an implicit stratification by minority group for the second-stage sample. A systematic probability proportionate to size sample of segments was selected from the sorted frames. A total of 1,959 segments were selected for the national NAAL sample.

For the SAAL states, a total of 861 segments were selected with probability proportional to MOS, with no oversampling of minority groups. The MOS and probability of selection are defined in table 7-8. The segments were selected independently within each state. For overlapping PSUs, the segments were selected independently for the NAAL and the SAAL. Two segments were selected for both the national NAAL sample and a SAAL sample. Therefore, a total of 2,818 unique segments were selected in the combined sample.

### **7.2.3.3 Selection of Housing Units**

The third stage of sampling involved selecting housing units from the frame of addresses in each segment, prepared after the listing operations were complete. After addresses had been selected, the interviewers contacted and screened households to determine whether they included any eligible respondent. Then, in low-minority segments, any household with at least one eligible person was included in the sample. In high-minority segments, which were oversampled, all minority households with at least one eligible person were retained in the sample, but only one-third of the nonminority households (with at least one eligible person) were included in the sample. This subsampling was done so that across all sampled segments, the resulting probabilities for nonminority households were equal under this scheme, which improves the precision of estimates for nonminorities since the variation in sampling probabilities was removed. In the SAAL samples, all households with at least one eligible person were retained in the sample (the SAAL samples did not include any oversampling of minority groups).

The national NAAL sampling approach ensured that the following conditions were met:

- At least one person was selected in each sampled household containing eligible persons 16 years or older, except that in nonminority households in high-minority segments, one-third of households were retained in the sample.
- In households of equal size
  - the probabilities of selection for persons in nonminority households were equal; and
  - the probabilities of selection for persons in minority households in high-minority segments were three times the selection probability of persons in all other selected households.

The housing unit probabilities of selection are presented in table 7-8. A reserve sample of about 5 percent the size of the main sample was selected randomly and set aside to be used in case of a shortfall in the sample. The reserve sample was not released. Housing unit sample sizes are provided in table 7-2.

The following quality control checks are examples of steps taken to ensure the high quality of the samples selected within the sampled segments:

- For the purpose of monitoring the listing operation, a range was generated for the expected number of housing units in the segment (+ 10 percent of the 2000 census count). Whenever the number of listed housing units fell outside the range, the lister called the data collection contractor and provided the reasons for the discrepancy.
- The within-segment sampling rates were applied to the count of listed housing units in the segment. This number was compared with the housing units subsampled from the field listing. This approach provided a check on the subsampling operations.
- From the listings, the address and identification number were keyed and verified. The keyed listings were checked by home office staff against the listing sheets. Any necessary corrections were made before household folders and assignment logs were produced.
- As the survey got under way, regular quality checks on the age, gender, and race distributions of the sample persons were made by comparing actual with expected distributions for the PSU.

The sample assignments, which specify the households and within-household subsampling, were checked by applying the algorithm used to generate them to the segment estimates of persons and households for minority and nonminority sample cases.

#### **7.2.3.4 Selection of Persons**

After selection, the addresses of the sampled housing units were loaded into the CAPI system. The within-household sampling was conducted by randomly selecting one adult from households with three or fewer eligible persons or two adults in households with four or more eligible persons. The random selection algorithm was programmed into the CAPI system, and the selected person(s) was displayed on the screen. The random selection of respondents was accomplished by assigning random numbers to each eligible person in the household and selecting the person(s) with the smallest (or the two smallest) random number(s). The selection of two adults in households with four or more eligible persons prevented a substantial increase in variances owing to high sampling weights, which would have resulted if the survey had selected only one person in households with large numbers of eligible adults. Sizes of the person samples are provided in table 7-2.

Most residents of college dormitories were expected to be available for interview at their family homes because the data collection period included the spring and summer breaks. However, if it was not possible to reach students at their family homes, arrangements were made to interview as many as feasible in their dormitories.

#### **7.2.3.5 Procedures for Selecting Missed Structures and Hidden Housing Units**

The missed structure and hidden housing unit procedures were developed to correct for any undercoverage that occurred during the listing operation. Procedures were implemented during data collection to handle any housing units identified through the hidden housing unit and missed structure procedures. The hidden housing unit procedure looked for housing units within a structure not included during the listing operation. If five or more hidden housing units were found, the statistician and the field director determined whether a sample of the hidden housing units had to be selected. Any sampling reduced the amount of interview work and clustering within the segments; however, it also increased the sampling error because of the unequal probabilities of selection resulting from the subsampling procedure.

For the missed structure procedure, interviewers looked for entire structures missed during the listing operation within a subsample of segments. The subsample of segments designated for the quality check was selected at a rate such that the inclusion of all units found retained the self-weighting feature of the sample stratum. If more than five missed structures were found in the segment, the statistician and the field director determined whether a sample of the missed structures had to be selected. As with the missed housing unit procedure, any sampling reduced the amount of interview work and clustering within the

segments; however, it also increased the sampling error because of the resulting unequal probabilities of selection.

In the national NAAL sample, 288 housing units were added through the missed structure procedure, and 153 housing units were added through the hidden housing unit procedure. Table 7-11 provides the number of housing units added by the missed structure and hidden housing unit procedures for the NAAL sample and each of the SAAL states. In total, 1.8 percent of the NAAL/SAAL combined sample consisted of housing units added through the missed structure and hidden housing unit procedures.

**Table 7-11. Number of housing units added by the missed structure and hidden housing unit procedures: 2003**

Sample	Total	Housing units added through missed structure procedure	Housing units added through hidden housing unit procedure
Total combined NAAL/SAAL	662	455	207
NAAL	441	288	153
SAAL	221	167	54
Kentucky	40	35	5
Maryland	19	10	9
Massachusetts	43	23	20
Missouri	8	7	1
New York	18	0	18
Oklahoma	93	92	1

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

### 7.2.3.6 Overall Probabilities of Selection

Table 7-8 provides the selection probabilities at each stage of sampling. The conditional probability in column 2 shows the selection probability at a particular stage of sampling, and the overall probability in column 3 shows the multiplicative overall selection probability across all previous stages. Therefore, the overall probability in the last row in table 7-8 is the overall probability of selecting a person into the NAAL or SAAL samples.

Hence, an explicit expression for the overall probability of selecting a person into the national NAAL sample is given by the product of the conditional probabilities from each selection stage (given that the reserve sample was not released):

$$P_{ijkl} = P_{hi} CP_{ij} CP_{ij(chunk)} CP_{ijk} CP_{ijk(mdu)} CP_{ijkl},$$

which, for NAAL, reduces to

$$P_{ijkl} = \frac{\bar{d}}{I_{NAAL}} \frac{mhu_{ijk}}{MHU_{ijk}} \frac{m_{ijk}}{M_{ijk}} \times 0.95$$

and, for SAAL, reduces to

$$P_{ijkl} = \frac{\bar{d}}{I_{SAAL}} \frac{mhu_{ijk}}{MHU_{ijk}} \frac{m_{ijk}}{M_{ijk}} \times 0.95$$

Where,

$mhu_{ijk}$  = number of selected missed housing units associated with housing unit  $k$ , within segment  $j$  of PSU  $i$ ,

$MHU_{ijk}$  = number of missed housing units found that were associated with housing unit  $k$ , within segment  $j$  of PSU  $i$ ,

$m_{ijk}$  = number of eligible persons selected within housing unit  $k$ , within segment  $j$  of PSU  $i$ ,

$M_{ijk}$  = number of eligible persons within housing unit  $k$ , within segment  $j$  of PSU  $i$ ,

$I_{NAAL}$  is the sampling interval for the selection of the NAAL segments, and

$I_{SAAL}$  is the sampling interval for the selection of the SAAL segments.

## **7.3 PRISON SAMPLE**

This section describes the design and selection of the state and federal inmate sample for the 2003 NAAL Prison Study. A two-stage sample was used to select inmates. At the first stage, 114 prisons were selected from the frame, with probabilities proportionate to an MOS. At the second sampling stage, an average of about 12 inmates were selected from the participating sampled facilities.

Section 7.2.1 describes the sampling frames for the study. Procedures used to select the sample of prisons and to select inmates within those facilities are described in sections 7.2.2 and 7.2.3, respectively.

### **7.3.1 Sampling Frames for the Prison Study**

The sampling frames for the Prison Study are discussed below, including an overview of the frames and data sources used to create the sampling frames.

#### **7.3.1.1 Overview of Prison Study Sampling Frames**

The target population consisted of inmates 16 years and older from state, federal, and private prisons in the United States. The sampling frame was created primarily from two data sources: the Bureau of Justice Statistics 2000 Census of State and Federal Adult Correctional Facilities (referred to in the following text as the Prison Census) and the 2003 Directory of Correctional Facilities of the American Correctional Association (ACA).

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

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

The Prison Study sample also included private facilities housing prisoners under exclusive contract to state governments and the Federal Bureau of Prisons.

Specifically excluded from the NAAL Prison Study were

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

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

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

### **7.3.1.2 Data Sources Used to Create Sampling Frames for the Prison Study**

The Bureau of Justice Statistics 2000 Prison Census included more than 1,600 facilities meeting the criteria provided in the previous section. The Prison Census data included facility addresses, capacity, inmate population, and security level, all of which were important information for sampling and data collection.

The 2003 ACA directory contained an updated list of more than 6,000 adult and juvenile state correctional departments, institutions, programs, and probation and parole/aftercare services. The directory also included updated inmate population figures, security level, and gender of the inmates, which were all helpful for sample design purposes.

The Prison Census list of facilities was compared with the ACA directory list to arrive at a sampling frame of prisons eligible for the study. After comparing the ACA and Prison Census information, project statisticians needed clarification for cases with unknown eligibility status. The data collection contractor called each state’s department of corrections and the Federal Bureau of Correctional Facilities to verify that the facilities were eligible for the study or to retrieve missing sampling information. Additionally, the number of cases in question was greatly reduced by obtaining information from various corrections-related websites.

Before sample selection, much work was done to separate work camps, annexes, satellites, and boot camps from their main facility. The sources used for this separation were the ACA directory, telephone calls, and websites. Table 7-12 shows the results of the frame creation operation.

The facilities were selected in late 2003, and inmates were selected and assessed in early 2004. The selection procedures are detailed in sections 7.2.2 and 7.2.3.

**Table 7-12. Summary of data sources used to create the prison sampling frame: 2003**

Source <sup>1</sup>	Count	Percent
Total	1,837	100.0
Census	1,559	84.9
American Correctional Association	92	5.0
Website	1	0.1
Separated from main facility on basis of ...		
Telephone call	7	0.4
Website	51	2.8
American Correctional Association Directory	127	6.9

<sup>1</sup> 2000 Census of State and Federal Adult Correctional Facilities (Census) and 2003 American Correctional Association Directory (ACA).

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

### 7.3.2 Selection of Prison Sample

The first-stage sampling units (or PSUs) were state or federal adult correctional facilities. The list of prisons was sorted before sample selection, which implicitly stratified the facilities (as described below), resulting in lower sampling variation than would be achieved with a simple random sample of facilities. The prisons were systematically selected from the sampling frame with probabilities proportional to the number of inmates in the facility.

The probability of facility  $i$  being selected is given by

$$P_i = \frac{a\hat{x}_i}{\sum_{i=1}^A \hat{x}_i},$$

where  $A$  indicates the number of prisons on the sampling frame,  $a$  indicates the number of prisons selected for the sample, and  $\hat{x}_i$  indicates the estimated number of inmates in facility  $i$  as it appeared on the sampling frame.

To determine the best sort order for the sampling frame, the data collection contractor conducted regressions of three literacy measures (prose, document, and quantitative) from the 1992 NALS, using census region and security level as the independent variables (the only variables on the current sampling frame that were also on the 1992 NALS public use file). The results showed that census region was a significant variable in explaining variation in literacy among inmates. Therefore, the facilities on the frame were ordered by census region first, followed by security level (supermaximum/maximum, medium, minimum, or other), type (federal, state, or private), and the number of inmates in the facility.

The frame was sorted in a serpentine fashion, with census region sorted first, in ascending order. Within the first level of census region, security level was sorted in ascending order. Within the second level of census region, security level was sorted in descending order. The same pattern continued for all of the sort variables. This process resulted in a list of prisons in which like units were adjacent to each other more often than in the traditional sort order.

A sample of 114 prisons was selected, allowing for prison nonresponse or ineligibility (e.g., closed). Table 7-13 provides the distribution of the prison sample by cooperation status and census region.

**Table 7-13. Distribution of selected prisons, by cooperation status and census region: 2003**

Prison-level characteristic	Total prisons sampled	Cooperation status		
		Participants	Refusals	Ineligibles
Total	114	107	3	4
Census region <sup>1</sup>				
Northeast	17	16	0	1
Midwest	22	19	3	0
South	50	48	0	2
West	25	24	0	1

<sup>1</sup> Northeast Region = ME, NH, VT, MA, RI, CT, NY, NJ, PA; Midwest Region = OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS; South Region = DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX; West Region = MT, ID, WY, CO, NM, AZ, UT, NV, WA, OR, CA, AK, HA.

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

### 7.3.3 Selection of Inmates Within Facilities

The second-stage units consisted of inmates selected within a sampled prison. Inmates were selected with a probability inversely proportional to the prison's population size so that the product of the first- and second-stage selection probabilities would be constant. In practice, the number of sampled inmates varied within prisons because of differences between the anticipated and actual sizes of the inmate populations and also because of constraints on the sample size per prison. The sample design was intended to provide a constant overall probability of selection across all inmates.

The conditional probability of inmate  $j$  being selected in prison  $i$  is given by

$$CP_{ij} = \frac{b_i}{\hat{x}'_i}$$

Where

$b_i$  = the expected number of inmates to be selected in prison  $i$  and

$\hat{x}'_i$  = the updated inmate population of prison  $i$ , obtained through a telephone call to the facility after its selection into the sample.

The expected number of inmates to be selected in prison  $i$  was calculated as

$$b_i = \frac{R_i}{k} b$$

where

$b = 12.12$ , the average inmate sample size inflated for anticipated nonresponse;

$$R_i = \frac{\hat{x}'_i}{\hat{x}_i}; \text{ and}$$

$$k = \frac{a'}{\sum_{i=1}^{a'} \frac{1}{R_i}}$$

and where

$a' = 107$ , the number of participating prisons.

The expression for  $k$  is the harmonic mean of the  $R_i$  values. Note that  $k$  is equal to 1 if  $\hat{x}'_i = \hat{x}_i$  for all sampled prisons (i.e., the number of inmates on the frame is equal to the number of inmates in the prison as determined through telephone contact). If all  $\hat{x}'_i > \hat{x}_i$ , then  $k$  will be greater than 1, and  $b_i$  will tend to be an average of size  $b$ .

Substituting the expression of  $b_i$  in the formula for the conditional probability,  $CP_{ij}$  gives

$$CP_{ij} = \frac{R_i b}{k \hat{x}'_i} = \frac{\hat{x}'_i}{\hat{x}_i} \frac{\sum_{i=1}^{a'} \frac{1}{R_i}}{a'} \frac{b}{\hat{x}'_i} = \frac{b \sum_{i=1}^{a'} \frac{1}{R_i}}{a' \hat{x}_i}.$$

The overall selection probability of an inmate is thus

$$P_{ij} = P_i CP_{ij} = \frac{a \hat{x}_i}{\sum_{i=1}^A \hat{x}_i} \frac{b \sum_{i=1}^{a'} \frac{1}{R_i}}{a' \hat{x}_i} = \frac{ab \sum_{i=1}^{a'} \frac{1}{R_i}}{a' \sum_{i=1}^A \hat{x}_i}.$$

Note that  $P_{ij}$  is constant across all inmates.

An upper bound of 16 inmates was set to constrain the size of the inmate sample per prison. This upper bound was dictated by the practical limitations of interviewing a large number of inmates in each

prison. If a prison's expected sample size exceeded the upper bound, it was truncated to the upper bound, and the sample sizes for the other prisons were inflated to yield the total expected inmate sample size. This iterative process was continued until there were no prisons with an expected inmate sample size greater than the upper bound. In addition, a lower bound of 9 inmates was set to justify the cost of traveling to prisons with a small number of interviewers.

Inmates in each prison were selected from a list of inmates occupying a bed the previous night. The interviewers received forms and instructions to follow when sampling inmates from the lists.

The interviewers had a laptop computer preprogrammed with a sampling algorithm. The statisticians assigned both the random number and the sampling interval to the prison before the fieldwork began; these values were preloaded into the sampling algorithm. The facility name, location, security level, type (federal, state, private), and gender composition (male only, female only, mixed) were also loaded on the laptop. The interviewers were required to verify all sampling information because it was also used in the sample weighting process.

The algorithm required that the interviewer enter the number of inmates on the list. After the number of inmates had been entered, the sampling algorithm compared the value with a preloaded acceptable range. In general, the acceptable ranges were within 10 percent of the expected inmate population. If the number of inmates fell outside the acceptable range, a message appeared on the laptop instructing the interviewer to contact the home office to receive a new sampling rate. After the interviewer entered the appropriate sampling rate, the laptop displayed the number of inmates to be sampled and the selected line numbers. The interviewer circled these line numbers on the list, and those inmates were selected.

Tables 7-14 through 7-16 show the background questionnaire and exercise sample counts for the inmate sample by prison and inmate characteristics. Weighted response rates are provided in chapter 11. Another component of the NAAL assessment was the Adult Literacy Supplemental Assessment (ALSA). A person took an ALSA assessment if he or she did not pass the core assessment. There were 29 inmates who took the ALSA assessment.

**Table 7-14. Background questionnaire sample counts by cooperation status and prison characteristics: 2003**

Prison-level characteristic	Actual total inmates sampled	Complete <sup>1</sup>	Nonresponse
Total	1,298	1,173	125
Prison type			
Federal	146	136	10
State/private	1,152	1037	115
Security level			
Supermaximum/ maximum	311	263	48
Medium	583	536	47
Minimum/other	404	374	30
Gender composition of prison			
Male	1,206	1,087	119
Female only /mixed gender composition	92	86	6
Census region <sup>2</sup>			
Northeast	183	159	24
Midwest	237	219	18
South	589	556	33
West	289	239	50

<sup>1</sup> Completed background questionnaires included cases that were not complete due to language problems and mental disabilities. These cases were considered a “success” in data collection since race/ethnicity, age, and gender were collected, as well as good information (language problem or mental disability) as to their English literacy skills.

<sup>2</sup> Northeast Region = ME, NH, VT, MA, RI, CT, NY, NJ, PA; Midwest Region = OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS; South Region = DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX; West Region = MT, ID, WY, CO, NM, AZ, UT, NV, WA, OR, CA, AK, HA.

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

**Table 7-15. Exercise sample counts by cooperation status and prison characteristics: 2003**

Prison-level characteristic	Total exercises attempted	Complete <sup>1</sup>	Nonresponse
Total	1,161	1,147	14
Prison type			
Federal	136	133	3
State/private	1,025	1,014	11
Security level			
Supermaximum/maximum	259	255	4
Medium	531	528	3
Minimum/other	371	364	7
Gender composition of prison			
Male	1,076	1,062	14
Female only/mixed gender composition	85	85	0
Census region <sup>2</sup>			
Northeast	154	152	2
Midwest	218	216	2
South	551	545	6
West	238	234	4

<sup>1</sup> Includes exercises coded as complete, reading/writing barrier, language problem, mental disability, or physical disability, as well as those coded as partial completes for the following reasons: reading/writing barrier, language problem, mental disability, or physical disability.

<sup>2</sup> Northeast Region = ME, NH, VT, MA, RI, CT, NY, NJ, PA; Midwest Region = OH, IN, IL, MI, WI, MN, IA, MO, ND, SD, NE, KS; South Region = DE, MD, DC, VA, WV, NC, SC, GA, FL, KY, TN, AL, MS, AR, LA, OK, TX; West Region = MT, ID, WY, CO, NM, AZ, UT, NV, WA, OR, CA, AK, HA.

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

**Table 7-16. Exercise sample counts by cooperation status and background questionnaire variables in the Prison Study, by inmate characteristic: 2003**

Inmate characteristic	Total exercises attempted	Complete <sup>1</sup>	Nonresponse
Total	1,161	1,147	14
Age			
16–29	388	386	2
30–49	659	647	12
50+	114	114	0
Gender			
Male	1,086	1,072	14
Female	75	75	0
Race			
Hispanic	223	222	1
Non-Hispanic Black only	491	482	9
Other	447	443	4
Education			
Less than high school	646	638	8
High school	347	342	5
More than high school	168	167	1

<sup>1</sup> Includes exercises coded as complete, reading/writing barrier, language problem, mental disability, or physical disability, as well as those coded as partial completes for the following reasons: reading/writing barrier, language problem, mental disability, or physical disability.

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

#### 7.4 ASSIGNMENT OF BOOKLETS TO RESPONDENTS

The NAAL assessment used 26 types of booklet to measure prose, document, and quantitative scales. The booklet types were assigned randomly to the housing units selected for the NAAL and SAAL samples. Before data collection, the selected housing units were sorted by PSU, segment, and their geographic sequence within their segment. Booklet types 1 through 26 were assigned to each housing unit by picking a random number  $l$  between 1 and 26 and assigning it to the first housing unit record. Booklet types for the remaining housing units were assigned sequentially  $l+1, l+2, \dots, 26, 1, 2, \dots, 26, 1, 2, \dots$ . The sequential numbers became the value of the booklet type assigned.

Each interviewer carried a spare bundle of booklets. The booklet types within each spare bundle were sorted with a random start between 1 and 26. If two persons were selected in a household, the top booklet from the spare bundle was administered to the second sampled person. Under the assumption of

random nonresponse, the booklet type assignment process was expected to result in an equal distribution of booklet types across respondents.

A similar process was developed to randomly assign booklet types to inmates in the Prison Study. Each prison was assigned a random set of booklet types, and the interviewers took the bundle of booklets and distributed them systematically to each sampled inmate. The prison sample booklet assignment process was expected to result in an equal distribution of booklet types across inmates.

*This page is intentionally left blank.*