
WEB DOCUMENTATION FOR NAAL GENERAL AUDIENCE REPORT

Introduction

This document describes the sampling and data collection, weighting and variance estimation, scaling, and statistical tests used in the collection and analysis of data for the 2003 National Assessment of Adult Literacy. The document also includes definitions of the variables used to report literacy results for different population groups. More detailed information about these topics can be found in the forthcoming 2003 NAAL Technical Report.

Sampling and Data Collection

The 2003 National Assessment of Adult Literacy included two samples: (1) adults age 16 and older living in households and (2) inmates age 16 and older in federal and state prisons. Each sample was weighted to represent its share of the total population of the United States. The household and prison samples were combined to create a nationally representative sample of America's adults. Household data collection was conducted from March through December of 2003; prison data collection was conducted from March through July of 2004.

Sampling and Response Rate for the Household Sample

The 2003 National Assessment of Adult Literacy household sample included a nationally representative probability sample of 35,365 households. Of these 35,365 sampled households, 4,671 were either vacant or not a dwelling unit, resulting in a sample of 30,694 households. In total, 25,123 households completed the screener which was used to select assessment respondents. The final weighted screener response rate was 81.2 percent.

Based upon the screener data, 23,732 respondents aged 16 and older were selected to complete the background questionnaire and the assessment. Of the 23,732 household respondents selected, 18,186 completed the background questionnaire. Of the 5,546 respondents who did not complete the background questionnaire, 355 were unable to do so because of a literacy-related barrier; either the inability to communicate in English or Spanish (the two languages in which the background questionnaire was administered) or because of a mental disability. The final weighted response rate for the background questionnaire, including respondents who completed the background questionnaire and respondents who were unable to complete the background questionnaire because of language problems or a mental disability, was 75.6 percent.

Of the 18,186 adults age 16 or older who completed the background questionnaire, 17,178 completed at least one literacy task on each of the three scales—

prose, document, and quantitative—included on the assessment. An additional 504 were unable to answer at least one question on each of the three scales for literacy-related reasons. The final weighted response rate for the cognitive assessment, which includes respondents who answered at least one question on each scale or were unable to do so because of language problems or a mental disability, was 96.6 percent.

Cases were considered complete if the respondent completed the background questionnaire and at least one question on each of the three scales or if the respondent was unable to answer any questions for because of language issues (an inability to communicate in English or Spanish) or because of a mental disability. All other cases that did not include a complete screener, background questionnaire, and responses to at least one question on each of the three literacy scales were considered incomplete or missing. Based on these criteria, the overall weighted response rate for the household sample was 60.1 percent.

Eight hundred and fifty-nine respondents answered the background questionnaire but refused to complete the assessment for reasons other than language issues or a mental disability. For these respondents, answers to one assessment item on each scale were imputed based upon the answers from respondents with similar background characteristics. In addition, a wrong response on each scale was imputed for 65 respondents who started to answer the assessment but were unable to answer at least one question on each scale because of language issues or a mental disability. The final household reporting sample—including the imputed cases—consisted of 18,102 respondents. After including the cases for which responses to the assessment questions were imputed, the response rate for the household sample was 62.1 percent (18,102 cases with complete or imputed data and an additional 439 cases that had no assessment data because of language issues or a mental disability).

Sampling and Response Rate for the Prison Sample

In addition to the household sample, the 2003 assessment also included a nationally representative probability sample of inmates in federal and state prisons. One hundred fourteen prisons were selected to participate in the adult literacy assessment. Of the 114 prisons, 107 agreed to participate, three refused, and four were determined to be ineligible. The final weighted prison response rate was 97.3 percent.

Inmates age 16 and older were randomly selected from among the inmates in the 107 sampled prisons to complete the background questionnaire and assessment. Of the 1,298 selected inmates, 1,161 completed the background questionnaire. Of the 137 who did not complete the background questionnaire, 12 were unable to do so because of a literacy-related barrier; either the inability to communicate in English or Spanish (the two languages in which the background questionnaire was administered) or because of a mental disability. The final weighted response rate for the prison background questionnaire, including respondents who completed the background questionnaire and respondents who were unable to complete the background questionnaire because of language problems or a mental disability, was 90.6 percent.

Of the 1,161 inmates who completed the background questionnaire, 1,125 completed at least one literacy task on each of the three scales—prose, document, and

quantitative—included on the assessment. An additional 20 inmates were unable to answer at least one question on each of the three scales for literacy-related reasons. The final weighted response rate for the cognitive assessment, including respondents who answered at least one question on each scale or were unable to do so because of language problems or a mental disability, was 98.8 percent.

The same definition of a complete case was used for the prison sample as for the household sample and the same rules were followed for imputation (as discussed below). The final weighted response rate for the prison sample was 87.2 percent.

One response on each scale was imputed based upon background characteristics for 28 inmates who completed the background questionnaire but had incomplete or missing assessment data. A wrong response on each scale was imputed for the 3 inmates who started to answer the assessment but were unable to answer at least one question on each scale because of language issues or a mental disability. The final prison reporting sample—including the imputed cases—consisted of 1,156 respondents. After including the cases for which responses to the assessment questions were imputed, the weighted response rate for the prison sample was 88.3 percent (1,156 cases with complete or imputed data and an additional 17 cases that had no assessment data because of language issues or a mental disability).

Scaling

Item response theory (IRT) methods were used to estimate average scale scores in the assessment for the nation as a whole and for various subgroups of interest. IRT models the probability of answering a question correctly as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance (or some other trait) can be compared across groups, such as those defined by sex, race/ethnicity, or place of birth.

IRT models assume that an examinee's performance on each item reflects characteristics of the item and characteristics of the examinee. The most commonly used models characterize items by the level of proficiency that they require and the precision with which item performance reflects proficiency along that trait. Examinees are characterized by their proficiency. An examinee's performance on a particular item reflects item difficulty, his or her proficiency, and the effects of residual factors that are not correlated across items or individuals.

All models assume that all items on a scale measure a common ability or proficiency (e.g., prose literacy) and that the probability of a correct response on an item is uncorrelated with the probability of a correct response on another item, an assumption known as conditional independence. Items are measured in terms of their difficulty as well as their ability to discriminate among examinees of varying ability and the probability that examinees with low ability will obtain a correct response through guessing.

The assessment used two types of IRT models to estimate scale scores. The two-parameter logistic (2PL) model, which was used for dichotomous items (that is, items that are scored either right or wrong) takes the form,

$$P(x_{ij} = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + e^{-1.7a_i(\theta_j - b_i)}}$$

Where x_{ij} is the response of person j to item i , θ_j is the proficiency of person j , a_i is the *slope* or *discrimination* parameter for item i , and b_i is the *location* or *difficulty* parameter for item i .

For the partial credit items, the Graded Response Logistic (GRL) model was used. This model follows the 2PL model for the probability of a score of 1 (at least partially correct):

$$P(x_{ij} \geq 1 | \theta_j, a_i, b_{i1}) = \frac{1}{1 + e^{-1.7a_i(\theta_j - b_{i1})}}$$

it also follows the 2PL model for the probability of a score of 2 (completely correct):

$$P(x_{ij} = 2 | \theta_j, a_i, b_{i2}) = \frac{1}{1 + e^{-1.7a_i(\theta_j - b_{i2})}}$$

In the above equations, b_{i1} and b_{i2} are the step parameters corresponding to the response categories of partially or fully correct.

Linear transformation of the scales was used to link the 2003 assessment scales to the 1992 assessment scales by setting an origin and unit size to the reported scale means and standard deviations from the 1992 assessment.

Weighting and Variance Estimation

A complex sample design was used to select assessment respondents. The properties of a sample selected through a complex design could be very different from those of a simple random sample, in which every individual in the target population has an equal chance of selection and in which the observations from different sampled individuals can be considered to be statistically independent of one another. Therefore, the properties of the sample for the complex data collection design were taken into account during the analysis of the data. One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all respondents. All population and subpopulation characteristics based on the NAAL data used sampling weights in their estimation.

The statistics presented in this report are estimates of group and subgroup performance based on a sample of respondents, rather than the values that could be calculated if every person in the nation answered every question on the instrument. It is therefore important to have measures of the degree of uncertainty of the estimates. Accordingly, in addition to providing estimates of percentages of respondents and their average scale score, this report provides information about the uncertainty of each statistic.

Because the assessment uses clustered sampling, conventional formulas for estimating sampling variability that assume simple random sampling and hence independence of observations are inappropriate. Standard errors calculated as though the data had been collected from a simple random sample would generally underestimate sampling errors. For this reason, the NAAL assessment uses a Taylor series procedure based on the *sandwich estimator* to estimate standard errors (Binder 1983).

Statistical Testing

The statistical comparisons in this report were based on the t statistic. Statistical significance was determined by calculating a t-value for the difference between a pair of means, or proportions, and comparing this value to published tables of values at a certain critical level, called alpha level. The alpha level is an a priori statement of the probability of inferring that a difference exists when, in fact, it does not. The alpha level used in this report is .05, based on a two-tailed test. The formula used to compute the t statistic was as follows: $t = \frac{(P_1 - P_2)}{\sqrt{(se_1^2 + se_2^2)}}$, where P_1 and P_2 are the estimates to be compared and se_1 and se_2 are their corresponding standard errors.

Variable Definitions

Age

All respondents were asked to report their birth dates, and this information was used to calculate their age. Age groups reported are 16 to 18, 19 to 24, 25 to 39, 40 to 49, 50 to 64, and 65 and older. Age groups were selected to correspond to key life stages of adults, although stages will vary from person to person:

- 16-18: Completion of secondary education
- 19-24: College or job training
- 25-39: Early career
- 40-49: Mid career
- 50-64: Late career
- 65+: Retirement

Disability Status

All respondents were asked the following four questions:

1. Do you have any difficulty seeing the words and letters in ordinary newspaper print even when wearing glasses or contact lenses, if you usually wear them?
2. Do you have any difficulty hearing what is said in a normal conversation with another person even when using a hearing aid, if you usually wear one?
3. Have you ever been diagnosed or identified as having a learning disability?
4. Do you have any other health problem, impairment, or disability now that keeps you from participating fully in work, school, housework, or other activities?

The data were recoded into the following six categories:

- Vision problem only
- Hearing problem only
- Learning disability only
- Other disability only
- Multiple disabilities
- No disabilities

Respondents who reported single disabilities were assigned to the appropriate category (Vision problem only, Hearing problem only, Learning disability only, and Other disability only). Respondents who reported multiple disabilities were assigned to the “Multiple disabilities” category. Respondents who reported no disabilities were assigned to the “No disabilities” category.

Educational Attainment

All respondents were asked to indicate the highest level of education they completed in the United States. Respondents who went to school outside the U.S. were probed for the equivalent level of educational attainment. The following options were provided:

- Still in high school
- Less than high school
- Some high school

-
- GED or high school equivalency
 - High school graduate
 - Attended a vocational, trade, or business school after high school
 - College: less than 2 years
 - College: associate's degree (A.A.)
 - College: 2 or more years, no degree
 - College graduate (B.A. or B.S.)
 - Postgraduate, no degree
 - Postgraduate degree (M.S., M.A., Ph.D., M.D., etc.)

Respondents who reported less than high school or some high school were asked how many years of education they completed. For certain analyses, some of these groups were collapsed. For example, respondents who had completed postgraduate studies but had not received a degree were generally combined with adults who completed a postgraduate degree.

Labor Force Participation

Household respondents were asked to report what they were doing one week before the assessment was administered:

1. working at a full time job for pay or profit (35 hours or more)
2. working two or more part time jobs for pay, totaling 35 or more hours
3. working for pay or profit part time (1 to 35 hours)
4. unemployed, laid off, or looking for work
5. with a job but not at work
6. with a job but on family leave (maternity or paternity leave)
7. in school
8. keeping house
9. retired
10. doing volunteer work

For analysis, respondents were divided into four groups: adults working full time (or working two or more part time jobs); those working part time; those unemployed, laid off, or looking for work; and those not in the labor force. Adults in categories 1 and 2 were counted as being employed full time; those in category 3 were counted as being employed part time; those in category 4 were counted as unemployed; those in categories 5 and 6 were counted as not at work (and therefore omitted from the labor force analyses); and those in categories 7 through 10 were counted as not in the labor force.

Language Spoken Before Starting School

All respondents were asked which language or languages they learned to speak before starting school. Their responses were then used to divide respondents into 5 groups: English only, English and Spanish, English and other language, Spanish only, and Other language(s). The English and Spanish category includes adults who spoke languages in addition to both English and Spanish. The Spanish category includes adults who spoke Spanish and additional non-English languages.

Race and Ethnicity

In 2003, all respondents were asked either two or three questions about their race and ethnicity. The first question asked them to indicate whether they were Hispanic or Latino. If a respondent answered that he or she was Hispanic or Latino, the respondent was asked to choose one or more of the following groups to describe his or her Hispanic origin:

- Mexican, Mexican American, or Chicano
- Puerto Rican or Puerto Rican American
- Cuban or Cuban America
- Central or South American
- Other Hispanic or Latino background

Next, all respondents, including those who indicated they were Hispanic or Latino, were asked to choose one or more of the following groups to describe themselves:

- White
- Black or African American
- Asian
- American Indian or Alaskan Native

-
- Native Hawaiian or other Pacific Islander

Individuals who responded “yes” to the first question were coded as Hispanic, regardless of their answer to the second question. Individuals who identified more than one group on the second question were coded as Multiracial. The interviewer recorded the race/ethnicity of respondents who refused to answer the question. Respondents of Native Hawaiian or Pacific Islander origin were grouped with those of Asian origin. The White, Black, and Hispanic groups are reported separately. For some analyses, detailed categories of Hispanic ethnicity were also reported.

In 1992, the race and ethnicity questions were somewhat different. Respondents were first asked to choose one race from among the following:

- White
- Black (African American)
- American Indian
- Alaskan Native
- Asian
- Other

Respondents were then asked if they were or of Spanish or Hispanic origin or descent. If a respondent indicated Spanish or Hispanic ancestry, he or she was asked to choose from among the same groups used in the 2003 survey to describe their Hispanic ethnicity.

Because respondents in 2003 were not offered an “Other” category to describe their race and respondents in 1992 were limited to choosing one race and ethnicity, caution should be exercised when comparing 1992 and 2003 results by race/ethnicity.

Sex

Interviewers recorded the sex of each respondent.

References

Binder, D.A. (1983). On the variances of asymptotically normal estimates for complex surveys. *International Statistical Review*, 51, 279-92.