

ESTIMATING LITERACY PROFICIENCIES WITH AND WITHOUT COGNITIVE DATA

Kentaro Yamamoto, Irwin Kirsch, Educational Testing Service
Kentaro Yamamoto, Educational Testing Service Princeton, NJ 08541

Key Words: Imputation, Non-response.

Abstract

In any survey of opinions, ability, or demographic characteristics, missing responses are always present for various reasons. The National Adult Literacy Survey (NALS) collected information as to why respondents broke off at any time during the interview. This information was used to determine how to treat missing cognitive data to make accurate inferences of English literacy proficiency of the entire adult population.

The NALS collected extensive background data as well as responses to the literacy exercises. The sample included 24,827 nationally representative housing units, of which 13% were vacant and 12% refused to participate, thus no detailed background information is available on this group. At the participating households, the interviewers asked screening questions to identify the number of age-eligible persons in the household. Depending on the number of eligible adults, one or more persons were selected to participate in NALS.

For the 26,091 persons who agreed to respond to the survey, extensive background information was collected before the cognitive exercises were presented. A great deal is known about these individuals, including their country of birth, language(s) spoken or read, age, highest level of education completed, current educational aspirations, labor market status, current occupation and wages, voting behaviors, and reading habits. After answering the background questions, however, 1,364 (5 percent) did not complete any cognitive exercises, and 1,630 (6 percent) responded to less than five exercises at least in one scale—the minimum number needed to estimate their proficiencies accurately. Omitting these individuals from the analyses would have resulted in overestimates of the literacy skills of the national population as a whole and particularly of certain subpopulations. Accordingly, special procedures were developed to estimate their literacy proficiencies.

1 Method for Treating Missing Cognitive Data

A common way to handle missing data is to ignore it. Many large-scale assessment programs such as the National Assessment of Educational Progress (NAEP), the National Educational Longitudinal Study (NELS), and the 1985 Young Adult Literacy Survey (YALS) excluded those who did not respond from their analyses. Ignoring missing data may be the only option if nothing is known about differences between the respondent and nonrespondent populations. Post stratified weighting is equivalent to this. This approach would be used based on the assumption that missing cases are missing

at random and that the remaining observed cases are representative of the target population. Randomly missing data rarely occur in real data collection, however.

When response rates vary across subpopulations, reporting the results based only on the sample of respondents can yield both biased and inaccurate proficiency distributions for some subpopulations. In the 1985 young adult survey, for example, most of the excluded individuals were Hispanic. Thus, survey results for the Hispanic subpopulation are based only on the scores of those Hispanic adults who read English and do not reflect the lower proficiencies of their peers who cannot read English.

At NCES's request, ETS implemented an approach to minimize distortions in the population proficiency estimates due to nonresponse. Accordingly, one of the objectives of the NALS field study was to investigate the characteristics of nonrespondents and to probe their reasons for not responding to the cognitive tasks. Using this information, it is possible to compare the characteristics of adults with missing cognitive data to those of adults who did respond to the literacy assessment items.

1.1 Why does Nonresponse Occur?

Answering the NALS background questions required no reading skills; questions were read to respondents by the assessment interviewer. A Spanish version of the background questionnaire and bilingual interviewers were available to assist individuals whose native language was not English. If the respondent did not answer a sufficient number of background questions, his or her case was considered incomplete. Such cases were never incorporated into the database.

Response rates on the background questionnaire vary according to the type of question. For example, questions asking country of origin (first question in the booklet) and gender (last question in the booklet) had nearly 100% response rates, indicating that most respondents attempted to complete the entire questionnaire. The rate of nonresponse was higher for certain types of questions, however—for example, questions about income and educational background.

1.1.1 Types of Nonresponse

There are three types of nonresponse with respect to the cognitive tasks: not presented, not attempted/not reached, and omitted.

Not presented. Due to the matrix sampling design used in the NALS, each respondent received only a fraction of the literacy exercises in the item pool. Hence, a particular pattern of not presented items occurs

for every respondent. Such patterns were appropriately ignored when calculating respondents' proficiencies.

Not attempted/not reached. In some cases, respondents were given tasks that they did not complete because they did not reach them. Not attempted/not reached items are found consecutively at the end of the blocks and are termed "consecutively missing responses," or CMR.

Omitted. An omitted response is a nonresponse to a task that was presented to a respondent but which the respondent chose not to perform or was unable to perform. For the NALS, omitted responses were treated as wrong, since most certainly a random response to an open-ended item would result in a wrong response.

There are varying degrees of nonresponse to the cognitive items. Some individuals answered no cognitive items, while others completed fewer than five items per scale. For the purposes of this discussion, any respondent who completed less than five items on any scale is considered an incomplete case. When adults did not complete the minimum number of tasks, they were asked to indicate their reasons for not responding, as described below. These reasons were subsequently categorized as related to literacy (i.e., individual was inferred to be unable to respond) or unrelated to literacy (i.e., individual was inferred to be unwilling to respond or gave no reason for not responding).

1.1.2 Non-Interview Reports (NIR)

NALS implemented a standardized procedure called non-interview report (NIR) that gathered information on reasons for nonresponse, such as

- a) non-English language
- b) physical or mental disability
- c) reading and/or writing difficulty
- d) respondent refused
- e) someone refused for respondent
- f) maximum calls completed without interview
- g) individual unavailable during testing period
- h) other reason

Table 1: Distribution of Cognitive Item Respondents and Nonrespondents

Number of completed cog. items	Relationship to literacy/Reasons for non-response	Percentage of adults	
0 to 4	related to literacy	5.7	
	non-English/reading dis.		5.1
	mental retardation, learning dis.		0.6
	not related to literacy	6.6	
	refusal		2.7
	physical dis.		2.0
	other/missing		1.9
5 or more		87.8	87.8
Total		100.0	100.0

Option a was followed by a more specific question about which language was spoken by the person. Option b was followed by a question asking the individual to indicate the specific disability: learning disability, mental or emotional condition, mental retardation, hearing impairment, visual impairment, speech or language impairment, physical disability, or other. Obviously, the interviewer is not a diagnostician of disability; such information was obtained from the interviewee or examinee assistant during the interview. However, no assistance was allowed during the administration of the literacy tasks. Table 1 shows the distribution of the NIR variable.

1.1.3 Reasons Unrelated to Literacy

Refusal. Some individuals (2.7 percent) refused to complete the cognitive portion of the assessment. For them no information is available about their performance on the cognitive items. The only information available about their literacy skills is from their answers to certain background questions, such as the highest level of education attained.

Physical disability. To answer the cognitive items, respondents had to be able to read materials and respond to tasks in writing without help. Individuals with physical disabilities such as visual impairment or lack of motor skills (2 percent of the total sample) would not have been able to produce written responses. Their reason for not responding was considered unrelated to literacy skills because they may have been able to respond successfully if the modes of presentation and/or response were not paper and pencil.

Other. About 1.9 percent of the total sample did not respond to the literacy tasks for one of the following reasons: interviewer exceeded maximum number of contacts without making an appointment, respondent unavailable, other unspecified reasons, or missing information. The major reason for missing information was due to difficulty in determining why the respondent did not complete the survey. Although some individuals with missing information might belong to one of following related to literacy-related categories, there was not enough information to make such determinations.

1.1.4 Reasons Related to Literacy

Language or reading difficulty. Some individuals (5.1 percent of the total sample) did not complete the assessment because they had difficulty reading in the English language. In most cases, this is because English was their second language. The assessment was not designed to investigate the nature of language difficulties in detail; rather, it is designed to identify language or reading difficulties not caused by physical disabilities. Nonresponse for these individuals may indicate inability to read and write in English but not necessarily in other languages.

Mental or learning disability. The general household population is likely to include more diverse

populations than those found in educational institutions. For example, mentally disabled or learning disabled populations are seldom included in the sampling frameworks of traditional school-based assessments such as NAEP. The NALS included every household member within a specified age range, however. Less than 1 percent (0.4) of the total sample did not respond to the survey because they (or others who knew them) identified themselves as having a mental or learning disability. In studies of the mentally disabled population, about 1 to 2 percent of the total population are reported to be mentally disabled. Such disabilities may be present at birth or may be related to aging, among other factors.

1.2 Why Imputation?

Multiple imputation comes from the plausible value methodology developed by Mislevy (1991), following the work by Rubin (1987). The method makes it possible to obtain unbiased estimates of proficiency distributions of subpopulations by incorporating the relationships between individual background variables and the posterior proficiency distribution. The multiple values are sampled from the posterior distribution directly instead of summarized into a point estimate.

The plausible values methodology was necessary for several reasons. The number of items in each NALS scale (12 to 15) is small, compared to the number of items in achievement tests geared to measure individual abilities, such as SAT and GED, even considering that the NALS items are open ended rather than multiple choice. Consequently, the uncertainty of individual proficiency estimates is too large to be ignored altogether. Plausible values methodology retains the uncertainty of individual estimates in order to increase the accuracy of population estimates. While this approach does not provide the best estimates of individual scores, it is the best suited for population estimates. Detailed treatment of methodology and its application can be found in the 1992 NAEP Technical Report.

1.2.1 Five Imputation Methods Considered

Five imputation methods were developed to represent the range of implied causality of missing responses. Each of the five methods has a specific effect on the likelihood function. Regardless of which of the five methods is used, the prior distribution is the same. The five methods differ with respect to the assumed causality of consecutively missing responses (CMR) at the end of the test for those cases where the respondent answered fewer than five items on at least one scale. If consecutive or non-consecutive missing responses were found before a valid correct or incorrect response, such missing responses were treated as omitted items. No response made by any individual was ever overwritten by any method.

The five imputation methods were:

1. All CMR are treated as wrong responses.

2. All CMR are treated as wrong if the reason for not responding to the literacy tasks is related to literacy (e.g., language difficulty, mental disability, or reading difficulty not caused by physical disability). The CMR of nonrespondents indicating reasons not related to literacy (e.g., physical disability or refusal) are treated as not reached.
3. For individuals who indicate literacy-related reasons for not responding to the tasks and whose CMR start from one of the first five items, the CMR to the first five items are treated as wrong responses, and the remaining portion of the CMR are treated as not reached. The CMR of individuals who indicate a reason not related to literacy are treated as not reached.
4. Regardless of the reasons for nonresponse, if the CMR start within the first five items, the CMR within the first five items are treated as wrong responses, and CMR after the fifth item are treated as not reached, as is the case in method 3.
5. All CMR are treated as not reached.

The following table (table 2) represents the score assignment of the five methods. A score of 0 indicates wrong, and a score of 3 indicates not reached.

Table 2 Treatment of consecutively missing responses (CMR) in the five imputation methods.

Method	Reason for nonresponse/relationship to literacy	CMR for first five items	Remaining CMR
1	Not related	0	0
	Related	0	0
2	Not related	3	3
	Related	0	0
3	Not related	3	3
	Related	0	3
4	Not related	0	3
	Related	0	3
5	Not related	3	3
	Related	3	3

It should be noted that while methods 2 and 3 take into account the reason for nonresponse (related or unrelated to literacy), methods 1, 4, and 5 do not. Methods 1 and 5 are extremes. According to Method 1, physical disability and refusal are considered synonymous with inability to read. Most large scale assessments use method 5. Method 5 may not be so extreme if the survey population is fairly uniform—for example, an in-school population. The adult population as a whole is not uniform, however. Because it includes non-English speakers who reside in the United States, some of whom do not read English at all, method 5 ignores too much information. When a test is speeded the method 5 is a reasonable option, however, the survey was not timed; hence, it is not a speeded test. There were, of course, some practical time limits so that the inter-

viewer would not have to spend an unreasonable number of hours collecting information. On the other hand, respondents were encouraged to attempt all questions.

Including the nonresponding respondents with those who responded to more than five items per scale would result in somewhat lower overall proficiency means. Sensitivity of proficiency means against five methods were evaluated. For groups independent to the reasons for nonresponse (e.g., gender), mean proficiencies are affected by all five methods equally. However, groups that do interact with the reasons for nonresponse (e.g., level of education) could produce different results depending on which of the five methods is used.

The impact of the five methods on the grand total means by gender was nearly zero. Among less educated respondents, younger respondents, and respondents with smaller household income, however, proficiency means interacted with all five methods. Not all minority populations were affected in the same way by each of the five methods. The Black population showed very little variation among five methods, about the same variation as the White population (10 percent of a standard deviation). The Hispanic and Asian populations showed the strongest sensitivity to the five methods (about 30 percent of a standard deviation). This is because many of the nonresponding Hispanic and Asian persons indicated a literacy-related reason for not completing the assessment. It is also clear that excluding the nonresponse samples would severely overestimate the literacy proficiencies of some subpopulations, namely the Hispanic and Asian populations. To some degree, the proficiency means of less educated, younger, and poorer subpopulations would also be overestimated.

1.2.2 The Method Selected

Methods 2, 3, and 4 have a relatively invariant impact on mean proficiency. Further, method 2 does not treat nonresponses differentially based on an arbitrary minimum of five items. Accordingly, method 2 was selected as the most viable approach for including individuals without cognitive data into the sample. All individuals without cognitive data who are included in the analyses have fairly extensive background information.

It should be noted that all proficiency values were obtained based on two types of information: responses to the background questions and responses to the cognitive items. As an intermediate step, a functional relationship between these two sets of information was calculated for the total sample, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency is calculated from a posterior distribution that is the multiple of two functions: the conditional distribution of proficiency, given the pattern of background variables, and the likelihood function of proficiency, given the pattern of responses to the cognitive items.

Since exact matches of background responses are quite rare, NALS used more than 200 principal compo-

nents to summarize the background information, capturing more than 99 percent of the variance. This procedure was selected to minimize the bias that could result from the inclusion of nonresponding individuals. It implies that the relationships between background variables and proficiency are the same for responding and nonresponding individuals. With this assumption, not-reached coding has no bearing on the proficiency estimates given identical background information. In other words, the higher rate of refusals within certain subpopulations does not lower the mean proficiencies of those subpopulations, unless their responses to the background questions include a high percentage of false statements. Any subpopulation's heterogeneity is evidenced by the wide distribution of background variables. Samples that received a not-reached coding may reflect a particular distribution within a subpopulation. In such cases, the proficiency distribution for nonrespondents would differ from the distribution of respondents because of background variable differences alone.

2 Data Analysis

Procedures to incorporate proficiency values for nonresponding individuals were established based on the field study and were thus in place before the NALS data analyses were conducted. The selected procedure (earlier identified as "method 2") relies on the individual's self-reported reason for nonresponse. The validity of such reasons is an important concern, given that the method would fail without assurance from validity checks. Since respondents' anonymity and privacy had to be protected in this survey, there is no way to go back to collect further information about nonrespondents' reasons for not completing the assessment. Thus, only internal validity checks are possible.

Because the survey collected extensive background information, some of which should strongly relate to reasons for nonresponse, some specific predictions can be made that support the validity of the reasons for nonresponse. For example, individuals who were born outside of the United States tend to have learned English as adults, and tend to have lower literacy skills in English than those born in this country. Such individuals were more likely to indicate literacy-related reasons for nonresponse than to indicate reasons unrelated to literacy. Since more recent immigrants are largely Hispanic and Asian, the proportions of nonresponding individuals who gave literacy-related reasons for nonresponse are greater for these populations than for the White and Black populations. For those who gave reasons unrelated to literacy, however, there should not be an interaction with the country of birth and English as a second language variables. We would also expect older individuals to be more likely than younger adults to state physical disability as a reason for nonresponse. If these expectations are fulfilled, then the data provide supporting evidence of the validity of the reasons for nonresponse.

2.1 Reasons Not Related to Literacy

Most individuals who gave reasons not related to literacy for not completing the assessment are age 65 or older (78 percent); only 15.5 percent of adults in the total population are in this age group. Let p_1 and p_2 be the probabilities of giving reasons not related to literacy for the two age groups of younger than 65 and 65 and older, respectively. The ratio of p_2/p_1 can be derived as $p_2/p_1 = (.78 \cdot .845) / (.155 \cdot .22) = 19.3$. Thus, the over-65 population is 19 times more likely not to respond to cognitive items for reasons such as physical difficulties or refusals. This is consistent with the field test results. Two-thirds of the respondents over 65 chose a physical disability as their reason for nonresponse, compared with less than 20 percent of those under age 65. The interaction of age and types of reasons for nonresponse is very closely tied to physical disabilities.

If any of the non-literacy-related reasons were somewhat related to language skills, a major proxy for language skills (such as country of origin) should have a noticeable impact on the selection of non-literacy-related reasons. The idea is that persons with poor English skills might hesitate to show their own deficiency by refusing to participate in the study. On the other hand, if non-literacy-related reasons are fairly unrelated to language skill, the distributions for subgroups that differ in the proportion of immigrants should be similar to the distribution for the total population.

Among those who cited reasons for nonresponse that were unrelated to literacy, the percentage of adults born outside of the United States (10 percent) was nearly identical to the proportion of such adults in the main sample (11 percent). For subpopulations in which there has not been a great influx of recent immigrants, such as the African American and White populations, the percentages of foreign-born nonrespondents giving reasons unrelated to literacy (6.6 percent of African American, 4.2 percent of White nonrespondents) were very similar to the percentage of foreign-born respondents in the total sample (6 percent and 4 percent, respectively). For subpopulations with a high proportion of recent immigrants, a slightly higher percentage gave non-literacy-related reasons (50 percent vs. 46 percent for the Hispanic population, and 86 percent vs. 78 percent for the Asian/Pacific Islander population). The standard error for the Asian/Pacific Islander population is fairly large, because of the small sample size; thus, it is safe to conclude that country of birth and language skills are not related to choosing non-literacy-related reasons in this population.

2.2 Reasons Related to Literacy

If literacy-related reasons are truly language-related, then immigrants from non-English-speaking countries should be proportionally overrepresented in the samples who gave such reasons. Although 11 percent of the total population reported having been born outside the

United States, foreign-born adults represent 53 percent of the adults who gave literacy-related reasons for not completing the assessment. Because race/ethnicity interacts with the proportion of immigrants, it was also evident in the distribution of race/ethnicity. The distribution of those who cited literacy-related reasons was 15 percent for the Black population, 46 percent for the Hispanic population, 6 percent for the Asian/Pacific Islander population, and 30 percent for the White population; in the total population, the distribution was 11 percent of Black adults, 10 percent of Hispanic adults, 2 percent of Asian/Pacific Islander adults, and 76 percent of White adults. This clearly indicates that country of birth (a proxy for language usage) is strongly related to indicating literacy-related reasons for nonresponse.

Twenty-nine percent of the nonrespondents who gave literacy-related reasons were age 65 or older. The older population was about two times more likely than the younger population to give a literacy-related reason for nonresponse. Among those who gave this type of reason, the over-65 sample identified mental disability and reading difficulties more often than language problems. It should be noted that reading difficulty here did not include cases caused by physical disabilities. The type of literacy-related reason indicated most often by younger adults was non-English language. In addition the proportion of the two age groups (84.6 percent for age 16-64, 15.4 percent for age 65 and older.) who identified a non-English language as their reason is nearly identical to the proportion of these two age groups in the total population (84.5 percent for age 16-64, 15.5 percent for those age 65 or older).

Figure 1 shows the distribution of education levels among the four groups of nonrespondents and the total NALS population. Generally, nonrespondents tended to be less educated than adults in the total population. Having less education was more notable among those who chose literacy-related reasons for nonresponse. More than 50 percent had less than an eighth grade education. The majority of nonrespondents who had more than a secondary education were foreign born and age 65 and older.

Figure 1: Distribution of levels of education by reason for nonresponse and for total population^{2.3}

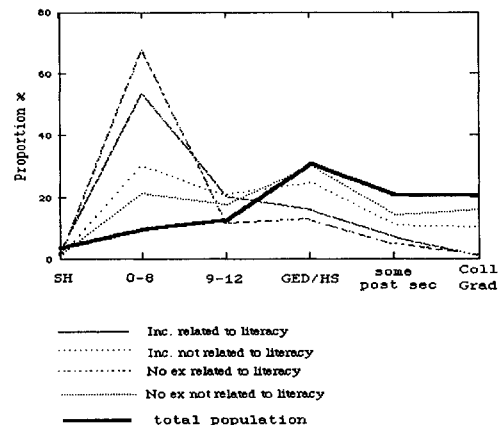


Table 3: Weighted proportion by type of nonresponse by age

Number of cognitive item responses	NIR category Reason relationship to literacy	Row percentage		Marginal proportion
		Age 16-64	Age 65+	
<5	Total	62	38	12.2
1 to 4	Total(related)	68	32	3.4
	non-English	86	14	1.3
	disability	56	44	2.1
	Total(unrelated)	47	53	2.9
0 items	refusal	75	25	0.4
	phy. disability	22	78	1.0
	other	66	34	1.4
	Total(related)	75	25	2.3
At least 5 items	non-English	84	16	1.6
	disability	54	46	0.7
	Total(unrelated)	59	41	3.6
	refusal	71	29	2.2
Total NALS sample	phy. disability	26	74	0.9
	other	78	22	0.3
At least 5 items		87	13	87.8
Total NALS sample		85	16	100.0

Table 4: Weighted proportion of foreign-born by type of nonresponse by race/ethnicity

Reason for nonresponse	Race/Ethnicity				
	Black	Hispanic	Asian	White	Total
Related to literacy	12.5	85.6	90.2	17.8	53
Not related to literacy	6.6	50.2	85.9	4.2	10
Total	6	46	78	4	11

Final Evaluation

In any survey of opinions, ability, or demographic characteristics, missing responses are always present. The most commonly practiced and least desirable way to treat missing data is to ignore it. This practice assumes that missing cases are missing at random and that the remaining observed samples are representative of the target population. This practice would yield both biased and inaccurate proficiency distributions for some subpopulations if response rates are different among subpopulations, and consequently for the total population, as well.

The procedure utilized in the NALS classified nonrespondents into two separate groups. One group can be thought of as a very unable population, including those who do not use English, those with extreme reading difficulties, and those with some type of cognitive disability. The other group did not respond for reasons that are not strongly related to literacy in English; these include adults who simply refused to respond to the assessment tasks, as well as those with physical disabilities.

Table 5: Weighted proportion by type of nonresponse by race/ethnicity

Nonresponse category/ Reason for nonresponse	Race/Ethnicity				Marginal percent Total
	Black	Hispanic	Asian	White	
Less than 5 items	16.1	31.7	21.2	8.6	12.2
1 to 5 items, related to literacy	6.3	11.8	10.5	1.6	3.4
	0.8	9.4	9.5	0.1	1.3
non-English disability	5.6	2.5	1.0	1.5	2.1
	4.6	2.5	2.6	2.6	2.9
1 to 5 items, not related to literacy	0.8	0.4	0.1	0.4	0.4
	1.4	0.3	1.0	1.0	1.0
refusal	2.4	1.8	1.6	1.2	1.4
	1.6	15.1	6.1	0.6	2.3
0 items, related to literacy	0.1	14.2	5.6	0.1	1.6
	1.5	0.9	0.6	0.5	0.7
0 items, not related to literacy	3.6	2.3	2.0	3.8	3.6
	1.9	1.4	1.0	2.4	2.2
refusal	1.3	0.6	0.4	0.9	0.9
	0.4	0.3	0.6	0.5	0.3
0 items, phy. disability	0.4	0.3	0.6	0.5	0.3
Other	0.4	0.3	0.6	0.5	0.3
At least 5 items	83.9	68.3	78.8	91.4	87.8
Total NALS	100	100	100	100	100

Responses to the background variables indicate that those who did not respond to the cognitive items for the literacy-related reasons were disproportionately likely to be foreign born, to have less than a high school education, to be Hispanic or Asian/Pacific Islander, and to be age 65 or older. These variables are known to relate strongly to English language proficiency and cognitive skills. Combined with other background information, there is strong evidence to support the notion that nonresponse to the cognitive items is not a random occurrence.

The above analysis assumes that adults' self-reported reasons for nonresponse are accurate and reliable. The accuracy of the NIR information is particularly important because of its impact on the proficiency distributions, particularly for some subpopulations. Background information largely affirmed nonrespondents' reasons for nonresponse. It is highly unlikely that this level of consistency could have occurred if the NIR information were erroneous. In future assessments of this kind, however, it might be advantageous to incorporate a system for monitoring the reliability of the NIR data.

References

- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.
- Mislevy, R.J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177-196.