

# NATES 2013: Nonresponse Bias Analysis Report

Evidence from a Nonresponse Follow-up Study

*This page is intentionally left blank.*

# NATES:2013 Nonresponse Bias Analysis Report: Evidence from a Nonresponse Follow-up Study

JANUARY 2017

**Michael Jackson**

**Rebecca Medway**

American Institutes for Research

**Sharon Boivin**

*Project Officer*

National Center for Education Statistics

NCES 2017-012  
U.S. DEPARTMENT OF EDUCATION



**U.S. Department of Education**

Arne Duncan  
*Secretary*

**Institute of Education Sciences**

Sue Betka  
*Acting Director*

**National Center for Education Statistics**

Peggy Carr  
*Acting Commissioner*

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

NCES, IES, U.S. Department of Education  
550 12<sup>th</sup> Street SW  
Washington, DC 20202

January 2017

The NCES homepage address is <http://nces.ed.gov>.

The NCES Publications and Products address is <http://nces.ed.gov/pubsearch>.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address as shown above.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-12-D-0002 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government

**Suggested Citation**

Jackson, M. and Medway, R. (2017). *NATES:2013 Nonresponse Bias Analysis Report: Evidence from a Nonresponse Follow-up Study* (NCES 2017-012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

**Content Contact**

Sharon Boivin  
(202) 245-7579  
[Sharon.Boivin@ed.gov](mailto:Sharon.Boivin@ed.gov)

# Executive Summary

## Introduction

Since 2009, a multi-agency research effort has sought to develop valid household survey measures of the attainment of non-degree, work-related educational credentials and training—including certifications/licenses, educational certificates, and work experience programs—among U.S. adults. This effort, coordinated by the Interagency Working Group on Expanded Measures of Enrollment and Attainment (GEMEnA), aims to address a key gap in federal data collection systems, which currently do not permit accurate and comprehensive measurement of these types of credentials. In support of this effort, the National Center for Education Statistics (NCES) administered three nationally representative pilot studies: the 2010 Adult Training and Education Survey (ATES), the 2013 National Adult Training and Education Survey (NATES), and the ATES topical component of the 2014 National Household Education Survey (NHES) Feasibility Study.<sup>1</sup> Informed by the results of these pilot studies, the first full-scale administration of the new ATES topical component of NHES took place in 2016. To date, individual items developed through this research effort have also been incorporated into several other federal surveys, including the U.S. Census Bureau’s Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS, sponsored by the Bureau of Labor Statistics and administered by the Census Bureau).<sup>2</sup>

The NATES pilot study, which is the focus of this report, tested the feasibility of using address-based sampling and a mailed questionnaire to collect detailed data on education, training, and credentials from U.S. adults. The target population for NATES was noninstitutionalized adults ages 16 through 65 who were no longer enrolled in high school. The NATES questionnaire included five key survey items asking about the following topics: attainment of certifications/licenses, attainment of educational certificates, current enrollment in college-level classes, completion of apprenticeships, and completion of other work-related trainings. Additional items requested detailed information about the characteristics of the reported credentials or classes and about respondents’ employment status and demographic characteristics. The mailed NATES questionnaire attained a response rate of 65.0 percent.

The primary purpose of the NATES pilot study was to determine whether a self-administered mailed questionnaire represented a feasible means of obtaining high-quality data from an adult target population. In particular, because prior adult-targeted NHES surveys have shown low response rates, the study sought to evaluate the extent to which nonresponse to the mailed questionnaire may lead to bias in the estimates. Broadly speaking, unit nonresponse bias is the deviation of a survey estimate from the true population value, attributable to the fact that not all sampled households or persons completed the survey

---

<sup>1</sup>The word “National” was removed from the survey title upon its incorporation into the National Household Education Surveys Program to avoid redundancy.

<sup>2</sup> The certification/license item was incorporated into the SIPP beginning with wave 13 of the 2008 SIPP (administered in 2013), and into the CPS beginning in 2015.

questionnaire. Unit nonresponse bias in a particular survey estimate is determined not only by the response rate to the survey, but also by differences between respondents and nonrespondents in terms of the characteristic of interest. Therefore, a “gold standard” nonresponse bias analysis requires respondents to be compared to nonrespondents in terms of characteristics known for both respondents and nonrespondents. In a typical household survey, this is not possible, because data on survey variables are not available for nonrespondents.

However, the NATES data collection was unique in that it incorporated a Nonresponse Follow-up Study (NRFU) that provides the data required for a nonresponse bias analysis (NRBA) comparing key estimates between respondents and nonrespondents to the mailed survey. In the NRFU, trained field interviewers administered a shortened, in-person version of the main NATES questionnaire to a random subsample of households that had not responded to the main, mailed survey. The NRFU questionnaire included key survey items asking whether respondents possessed a certification/license or certificate and whether they were currently enrolled in college-level classes. It also included a subset of the background demographic items from the NATES questionnaire. Altogether, data were successfully collected from approximately 83.1 percent of the NATES nonrespondents sampled for the NRFU. The availability of high-quality data from a representative subsample of nonrespondents allowed unit nonresponse bias in key NATES estimates—including estimates of the prevalence of key non-degree credentials—to be estimated directly.

## Purpose of This Report

This report presents methodological findings related to the research on nonresponse bias conducted for the NATES pilot study. The purpose of this research was as follows:

- to measure the extent of unit nonresponse bias in key NATES estimates, both prior to and after nonresponse adjustment, using data from the in-person NRFU;
- to determine whether methods that are available for nonresponse bias analysis in the absence of an NRFU, such as the use of auxiliary data and comparisons of selection-weighted to nonresponse-adjusted estimates, provide an accurate indication of the risk of bias and of the effectiveness of nonresponse adjustment in NATES;
- to evaluate two potential methods of leveraging data linked to the sampling frame to correct for unit nonresponse bias in future NATES administrations; and
- to investigate the possible effect of the NATES questionnaire design on the risk of item nonresponse bias.

The remainder of the Executive Summary provides an overview of the key findings and recommendations contained in the full report. The research effort described in this report was undertaken for questionnaire and procedural development purposes only. The information collected and published from this effort should not be used to generate or cite population estimates because the sample design and data collection procedures were not intended for that purpose, but rather to measure the potential for nonresponse bias in a household mail survey and to support the evaluation of questionnaire items.

## Findings

This section summarizes key findings from the analyses conducted for the report. It is divided into the following subsections, each corresponding to a chapter of the report:

- Unit nonresponse bias and the effectiveness of nonresponse adjustment (chapter 2);
  - After nonresponse adjustment, there was no statistically significant bias in any of the three key survey items included on both NATES and NRFU questionnaires: certifications/licenses, certificates, and college enrollment.
  - However, nonresponse bias was found and household-level nonresponse adjustments had little practical effect on several educational, age, racial, and income subgroups, for which bias remained statistically significant after adjustment.
- The usefulness of auxiliary data for correcting for nonresponse bias (chapter 3);
  - No consistent associations between available auxiliary variables and key survey variables were found, suggesting that the limited effectiveness of the nonresponse adjustment process is driven by low underlying correlations between these variables and key survey variables. This result also suggests that targeting specific subgroups defined by these variables would be unlikely to have a substantial impact on nonresponse bias.
  - However, households that were missing data for two or more auxiliary variables did differ with respect to some key estimates, suggesting that, if substantial improvements in unit response rates could be achieved among such households, nonresponse bias could be reduced.
- Questionnaire design and item nonresponse bias (chapter 4).
  - Item nonresponse rates were not trivial for many items that were dependent on skip patterns. Complex skip patterns in the NATES questionnaire appeared to increase the risk of item nonresponse bias especially for respondents who were less educated, were older (over 55), and/or spoke a language other than English at home.
  - Some of the characteristics that were found to be related to item nonresponse were also significantly associated with item responses. This implies that respondents and nonrespondents to these items differ from each other in ways that could result in a risk of item nonresponse bias in NATES items.

For a more detailed discussion of the analysis and results reported in each subsection, refer to the corresponding chapter of the report.

### ***Unit nonresponse bias and the effectiveness of nonresponse adjustment***

The data collected from the NATES nonrespondents sampled for the NRFU were used to estimate unit nonresponse bias for 26 proportions from nine survey items that were included in both the NATES and NRFU questionnaires. Unit nonresponse bias in each proportion was calculated as the difference between the proportion generated using only data from respondents to the mailed NATES (the “NATES-only proportions”) and the same proportion generated using the combined data from respondents to the mailed NATES and the in-person NRFU (the “NATES plus NRFU proportions”). When the NATES-only proportions were estimated using unadjusted selection weights (i.e., weights that account for differential sampling rates but not for nonresponse), bias in the estimated prevalence of certifications/licenses and certificates was found to be not statistically significant.

However, statistically significant bias was observed in the college class enrollment item: the NATES-only proportion underestimated college class enrollment by approximately 1.5 percentage points. Furthermore, five of the six demographic items included in both questionnaires showed statistically significant bias: the

NATES-only data overestimated the proportion with a bachelor's or graduate degree by about 2.3 percentage points; the proportion over the age of 55 by about 7.1 percentage points; the White proportion by about 5.1 percentage points; the non-Hispanic proportion by about 4.1 percentage points; and the proportion in the highest household income category (over \$75,000) by about 4.7 percentage points.

The NATES-only proportions were then re-estimated using nonresponse-adjusted weights. The nonresponse adjustments reduced the bias in the college class enrollment item from 1.5 percentage points to approximately 0.9 percentage points, which was no longer statistically significant. The bias in the certification/license and certificate items remained not statistically significant. The adjustments (which used auxiliary variables) led to limited changes in the demographic proportions (based on survey variables) and, therefore, in the amount of observed nonresponse bias in these proportions: statistically significant bias above 1 percentage point remained in 12 of the 20 categories of the educational attainment, age, race, Hispanic origin, and income items after nonresponse adjustment. Altogether, the nonresponse adjustments decreased the number of proportions showing statistically significant bias above 1 percentage point from 15 to 12 and reduced the absolute value of the estimated bias by an average of 19.9 percent over all 26 estimates. However, in a full-scale administration, the remaining bias in demographic proportions could be mitigated through the use of a poststratification or raking adjustment. These results therefore suggest that nonignorable bias in the NATES estimates that were evaluated using the NRFU is likely to be limited.

Additionally, for four NATES survey items that were not included in the NRFU questionnaire—including the remaining key survey items asking about the completion of apprenticeships and other work-related training—selection-weighted proportions were compared to nonresponse-adjusted proportions. For all but one of the items, the change in the proportion attributable to nonresponse adjustment was below 0.5 percentage points and not statistically significant. To the extent that bias existed in these estimates, the fact that the adjustments led to little change in the proportions suggests a minimal impact on the amount of nonresponse bias. However, because these items were not included in the NRFU questionnaire, it is not possible to determine whether these items showed significant bias prior to nonresponse adjustment.

Relative to the NRFU, other available methods of nonresponse bias analysis appear to either overestimate or underestimate the risk of bias in key survey estimates. In particular, comparisons of the distribution of auxiliary variables between respondents and the sample appear to overestimate the risk of bias prior to nonresponse adjustment, but also to overestimate the effectiveness of the adjustments at removing bias. Comparisons of selection-weighted and nonresponse-adjusted key estimates, in contrast, are likely to underestimate the magnitude of bias in the selection-weighted estimates.

### ***Usefulness of auxiliary data for correcting for nonresponse bias***

The NATES sampling frame included a number of auxiliary variables appended to it by the vendor from which the frame was purchased. These variables provided information on a range of household-level address and demographic characteristics. In principle, these auxiliary data have utility both for statistical and operational methods of correcting for nonresponse bias:

- The *statistical* nonresponse adjustment procedure used in NATES works by assigning sampled households to adjustment cells defined by auxiliary variables found to be associated with response propensity. The effectiveness of this procedure at correcting for nonresponse bias in a



particular estimate depends on the extent to which the auxiliary variables are associated with the characteristic of interest. Therefore, its effectiveness could be hindered by weak underlying associations between the auxiliary variables and the characteristics measured by the survey instrument and/or by high missing data rates for the auxiliary variables (the missing data rate in the NATES sample exceeded 40 percent for some variables). Additionally, the effectiveness of the nonresponse adjustment procedure could be affected by the measurement properties of the auxiliary data as well as the level of aggregation.

- *Operationally*, auxiliary data could be used to assign higher incentives, more intensive nonresponse follow-up, and/or other targeted strategies to particular subgroups that are less likely to respond. Because an increase in response rates would be expected to reduce nonresponse bias only if the additional respondents were likely to provide a different distribution of responses than existing respondents, the effectiveness of this approach again depends on whether auxiliary variables associated with response propensity are also associated with responses to key items.

Data from NATES and the NRFU were used to evaluate the likely utility of the auxiliary data for minimizing the risk of unit nonresponse bias in future ATES administrations through one or both of these methods.

As discussed above, the use of nonresponse-adjusted weights led to limited changes in the NATES estimates and, therefore, to limited changes in the amount of bias observed in these estimates. In order to further investigate the usefulness of the auxiliary data for statistically correcting for nonresponse, logistic regression was used to evaluate associations between the nine auxiliary variables used for the NATES nonresponse adjustments and the propensity to report one of the key credentials of interest. Among the households for which the auxiliary data on these characteristics were not missing, few statistically significant relationships were found. One variable, the age of the head of the household, was found to be associated with the propensity to report college class enrollment and work-related training. No other variables were significantly associated with the propensity to report more than one of the credentials of interest, although address route type was significantly associated with the reporting of a certification/license, home tenure was significantly associated with the reporting of college enrollment, and household income and the education of the head of the household were significantly associated with the reporting of work-related training. This finding suggests that the limited impact of the nonresponse adjustments on the NATES estimates was driven not only by high missing data rates for the auxiliary variables, but also by a limited association between those variables and credential reporting among households for which the variables were not missing. This, in turn, suggests that the efficacy of the nonresponse adjustment procedure would be unlikely to substantially improve even if data on these variables could be obtained for a larger proportion of sampled households.

The lack of consistent associations between auxiliary variables and key survey estimates also suggests that the operational approach of targeting specific subgroups defined by these variables would be unlikely to have a substantial impact on nonresponse bias in future ATES administrations. However, a potential variant of this approach, wherein households would be targeted based on whether they are missing data for a substantial number of these auxiliary variables, showed more promise. In general, response rates to NATES were found to increase as the number of auxiliary variables missing data decreased—for example, the response rate among households missing data for two auxiliary variables was 64.1 percent, compared to 66.7 percent among households missing data for a single auxiliary variable and 75.0 percent

among those with no missing auxiliary variables. Furthermore, a separate logistic regression analysis found a statistically significant association between the number of missing auxiliary variables and responses to the certification/license and college class enrollment items. This suggests that if substantial improvements in unit response rates could be achieved among households with missing data for multiple auxiliary variables, unit nonresponse bias in these particular estimates may be reduced. Therefore, the number of auxiliary variables with missing data may be usable as a targeting criterion in future administrations if missing auxiliary data remains correlated with responses to key survey items.

### ***Questionnaire design and item nonresponse bias***

In self-administered surveys such as NATES, certain questionnaire design features may cause respondents to fail to answer particular items. In principle, this can lead to an additional source of non-sampling bias, referred to as item nonresponse bias. In NATES, the missing data rates for the key survey items—those measuring the attainment of the credentials of interest—ranged from 1.3 to 2.8 percent, implying a negligible risk of item nonresponse bias. However, missing data rates were substantially higher for follow-up items related to the characteristics of reported credentials, many of which required respondents to navigate complex skip patterns. In order to evaluate the possible impact of these skip patterns on item nonresponse bias in NATES, bias analysis was conducted for 13 items that immediately followed skip directives and had missing data rates above 15 percent.

First, for each item, item respondents were compared to item nonrespondents in terms of four background characteristics: their reported educational attainment, age, household language, and employment status. For all 13 items, item nonrespondents were significantly more likely to have a high school degree or lower and/or significantly less likely to have a bachelor's degree or higher. For four items, item nonrespondents were significantly more likely to be above the age of 55. For three items, item nonrespondents were significantly more likely to speak a language other than English at home. These findings are consistent with the hypothesis that respondents who were less educated, were older, and/or spoke a language other than English at home may have had more difficulty following the skip patterns in the NATES questionnaire. For six items, there existed a statistically significant relationship between employment status and item response status; however, the direction of the relationship was inconsistent between the items, implying that the relationship between employment status (which was hypothesized to be a proxy for topic salience) and item response status is inconclusive.

In the second phase of the analysis, associations between the same four background characteristics and the actual response to each item were analyzed among those who responded to the item. For eight of the items analyzed, a characteristic that had previously been found to be significantly associated with item response status was also found to be significantly associated with the actual response to the item. Therefore, the differences between item respondents and nonrespondents in terms of observed characteristics—education, household language, age, and/or employment—may be suggestive of differences in terms of the characteristics measured by these items. In other words, this finding is suggestive of a risk of item nonresponse bias in these eight items.

## Recommendations

This section summarizes four recommendations that follow from the results of the analyses conducted for this report. Refer to chapter 5 for a more detailed discussion of each recommendation, as well as key limitations of the analyses.

### ***Recommendation 1: Target households with missing auxiliary data for multiple variables.***

While limited associations were found between the auxiliary data linked to the NATES sampling frame and responses to key survey items, the number of auxiliary variables for which a household was missing data were found to be significantly related both to response propensity and to several key items. Consequently, this study suggests that, if the auxiliary variables are used to target strategies such as higher incentives or more intensive nonresponse follow-up, the fact that a household is missing substantial amounts of auxiliary data is likely to be a more effective targeting criterion than information provided by the data itself. Additional research is necessary to determine whether such strategies can lead to increased response rates among these households and whether the association between missing auxiliary data and credential reporting remains consistent across ATES administrations.

### ***Recommendation 2: Use screener data for nonresponse adjustments and/or response rate targeting.***

The limited utility of commercial auxiliary data at mitigating nonresponse bias points to a need to exploit alternative sources of auxiliary data. Because future ATES administrations will be incorporated into the two-stage NHES design, one such source is data collected from household screeners. While screener data are unavailable for nonrespondents to the screener stage, they are available for all households sampled for the topical stage, including topical nonrespondents. In general, self-reported screener data are likely to be more accurate than commercially provided auxiliary data. For these reasons, information from household screeners is typically used to generate topical-level nonresponse adjustments in NHES. Consideration should also be given to using the information reported in the screener to target incentives or follow-up at the topical stage of data collection. While the use of screener data for nonresponse adjustment was not empirically evaluated in this report, this recommendation is rooted in the apparent disutility of the auxiliary data available in or linked to the sampling frame. Furthermore, it may be possible to improve the efficacy of screener data at reducing nonresponse bias by expanding the screener questionnaire to include items expected to correlate strongly with responses to key items at the topical stage. However, this consideration must be balanced with the need to minimize response burden at the screener stage.

### ***Recommendation 3: Explore alternative sources of auxiliary data on respondents and nonrespondents.***

The address-based sampling frames used for NHES typically include census tract and block group identifiers, which can be used to link sampled addresses to publicly available small area estimates from high-quality data collections such as the U.S. Census Bureau's American Community Survey. Therefore, appended block group- or tract-level demographic data represent a potential alternative source of high-quality auxiliary information for use in nonresponse adjustment and/or targeting. Given the relatively low cost of obtaining and appending census data, additional research should be conducted to assess their

potential efficacy at correcting nonresponse bias in future ATES administrations. A likely obstacle to this approach is that, insofar as household-level demographic characteristics correlate weakly with responses by individuals to survey items, data at a higher level of aggregation would intuitively be expected to show even less of an association. Consistent with this expectation, Biemer and Peytchev (2013) found that geocoded census data did not generate effective nonresponse adjustments for the National Comorbidity Survey Replication (a study sponsored by the National Institute of Mental Health).

***Recommendation 4: Minimize item nonresponse bias through simplified questionnaire design.***

Although item nonresponse bias cannot be easily evaluated, this report found evidence that suggests that the complex skip patterns in the NATES questionnaire increased the risk of item nonresponse bias (though the potentially greater difficulty of detailed follow-up questions may also be a driver of nonresponse). Therefore, the results of this study support the minimization of complex skip patterns in future self-administered questionnaires. In response to the high rates of missing data for some NATES items, as well as findings from cognitive testing, the skip patterns used in the 2014 NHES Feasibility Study and in NHES:2016 have been substantially simplified.

# Contents

EXECUTIVE SUMMARY .....	i
Introduction.....	i
Purpose of This Report .....	ii
Findings.....	ii
Recommendations.....	vii
LIST OF TABLES .....	xi
CHAPTER 1: INTRODUCTION .....	1
1.1 Need for Data .....	1
1.2 Topical Focus of the NATES Questionnaires.....	2
1.3 NATES Sampling Frame, Data Collection, Weighting, and Response Rates .....	3
1.4 Methodological Introduction.....	7
1.5 Key Research Questions and Structure of Report.....	9
CHAPTER 2: MEASURING UNIT NONRESPONSE BIAS .....	11
2.1 Evaluation of Bias Using Auxiliary Variables.....	11
2.2 Measurement of Bias Using the NRFU Follow-Up Sample .....	18
2.3 Evaluation of Nonresponse Adjustments for NATES-Only Items .....	24
2.4 Summary and Discussion of Results.....	25
CHAPTER 3: EVALUATION OF STRATEGIES FOR REDUCING UNIT NONRESPONSE BIAS....	27
3.1 Auxiliary Variables and Responses to Key Items.....	28
3.2 Availability of Auxiliary Variables and Responses to Key Items .....	37
3.3 Discussion of Results.....	42
CHAPTER 4: ANALYSIS OF ITEM NONRESPONSE BIAS .....	45
4.1 Differences Between Item Respondents and Nonrespondents.....	46
4.2 Analysis of Item Nonresponse Bias .....	48
4.3 Discussion of Results.....	49
CHAPTER 5: CONCLUSION.....	51
5.1 Summary of Findings.....	52
5.2 Limitations .....	56
5.3 Summary of Recommendations .....	58
REFERENCES .....	61

APPENDIX A: NONRESPONSE BIAS IN THE NONRESPONSE FOLLOW-UP SAMPLE.....	A-1
APPENDIX B: DETAILED TABLES FROM ITEM NONRESPONSE ANALYSIS .....	B-1
B.1 Differences Between Item Respondents and Item Nonrespondents.....	B-1
B.2 Response Distributions by Demographic Subgroups .....	B-14
APPENDIX C: CHARACTERISTICS OF HOUSEHOLDS MISSING AUXILIARY DATA .....	C-1
References.....	C-3
APPENDIX D: STANDARD ERROR TABLES.....	D-1
APPENDIX E: SAMPLING, WEIGHTING, AND ESTIMATION PROCEDURES .....	E-1
E.1 Sampling and Data Collection.....	E-1
E.2 Weighting .....	E-6
E.3 Variance Estimation .....	E-18
E.4 Response Rate Calculation .....	E-20
E.5 Combining NATES and NRFU Data .....	E-24
References.....	E-26
APPENDIX F: NATIONAL ADULT TRAINING AND EDUCATION SURVEY (NATES) QUESTIONNAIRE .....	F-1
APPENDIX G: NONRESPONSE FOLLOW-UP STUDY (NRFU) QUESTIONNAIRE.....	G-1

# List of Tables

Table 2.1.	Percentage distribution of NATES respondents and eligible sample, estimated bias, and percent relative bias, by weighting type and selected auxiliary variables: 2013.....	14
Table 2.2.	Percentage distribution of NATES plus NRFU respondents and NATES-only respondents, estimated bias, and percent relative bias, by weighting type and selected reported characteristics: 2013 .....	21
Table 2.3.	Percentage distribution of NATES respondents, by weighting type and selected reported characteristics: 2013 .....	24
Table 3.1.	Associations between auxiliary variables and reporting key estimates, by key survey item and characteristic: 2013 .....	31
Table 3.2.	Odds ratios of reporting a certification or license, by respondent sample and selected characteristics: 2013.....	32
Table 3.3.	Odds ratios of reporting a certificate, by respondent sample and selected characteristics: 2013.....	33
Table 3.4.	Odds ratios of reporting college enrollment, by respondent sample and selected characteristics: 2013.....	34
Table 3.5.	Odds ratios of reporting an apprenticeship, by selected characteristics: 2013.....	35
Table 3.6.	Odds ratios of reporting other work-related training, by selected characteristics: 2013.....	36
Table 3.7.	Percentage distribution of NATES eligible sample and weighted response rate to mailed NATES questionnaire, by number of key auxiliary variables missing for household: 2013 .....	38
Table 3.8.	Odds ratios of reporting a certification or license, by respondent sample and number of missing auxiliary variables: 2013 .....	40
Table 3.9.	Odds ratios of reporting a certificate, by respondent sample and number of missing auxiliary variables: 2013 .....	40
Table 3.10.	Odds ratios of reporting college enrollment, by respondent sample and number of missing auxiliary variables: 2013.....	41
Table 3.11.	Odds ratios of reporting an apprenticeship, by number of missing auxiliary variables: 2013.....	41
Table 3.12.	Odds ratios of reporting work-related training, by number of missing auxiliary variables: 2013 .....	41
Table 4.1.	Item missing rates and characteristics associated with item response status, by demographic characteristic and item: 2013.....	47
Table 4.2.	Characteristics associated with response status and response to item, by demographic characteristics and item: 2013 .....	49
Table A.1.	Percentage distribution of NRFU respondents and eligible sample, estimated bias, and percent relative bias, by selected auxiliary variables: 2013 .....	A-2
Table B.1.	Percentage distribution of respondents and nonrespondents to Q12 (activities to earn continuing education units), by selected reported characteristics: 2013 .....	B-1
Table B.2.	Percentage distribution of respondents and nonrespondents to Q18a/Q18b (certification/license for past or future job), by selected reported characteristics: 2013....	B-2

Table B.3.	Percentage distribution of respondents and nonrespondents to Q21 (subbaccalaureate certificate), by selected reported characteristics: 2013 .....	B-3
Table B.4.	Percentage distribution of respondents and nonrespondents to Q22 (year certificate was earned), by selected reported characteristics: 2013.....	B-4
Table B.5.	Percentage distribution of respondents and nonrespondents to Q28a/Q28b (certificate for past or future job), by selected reported characteristics: 2013 .....	B-5
Table B.6.	Percentage distribution of respondents and nonrespondents to Q30 (year of last apprenticeship), by selected reported characteristics: 2013 .....	B-6
Table B.7.	Percentage distribution of respondents and nonrespondents to Q34 (apprenticeship occupation), by selected reported characteristics: 2013.....	B-7
Table B.8.	Percentage distribution of respondents and nonrespondents to Q36 (classes to earn credential), by selected reported characteristics: 2013.....	B-8
Table B.9.	Percentage distribution of respondents and nonrespondents to Q37 (credential being earned), by selected reported characteristics: 2013 .....	B-9
Table B.10.	Percentage distribution of respondents and nonrespondents to Q39 (number of classes), by selected reported characteristics: 2013 .....	B-10
Table B.11.	Percentage distribution of respondents and nonrespondents to Q44 (employer reimbursing tuition), by selected reported characteristics: 2013.....	B-11
Table B.12.	Percentage distribution of respondents and nonrespondents to Q50 (training provided at no charge), by selected reported characteristics: 2013.....	B-12
Table B.13.	Percentage distribution of respondents and nonrespondents to Q78 (English ability), by selected reported characteristics: 2013 .....	B-13
Table B.14.	Percentage of NATES respondents reporting specified responses to Q12 (activities to earn continuing education units), by response and selected reported characteristics: 2013.....	B-14
Table B.15.	Percentage of NATES respondents reporting specified responses to Q18a/b (certification/license for past/future job), by response and selected reported characteristics: 2013.....	B-15
Table B.16.	Percentage of NATES respondents reporting specified responses to Q21 (type of certificate), by response and selected reported characteristics: 2013.....	B-16
Table B.17.	Percentage of NATES respondents reporting specified responses to Q22 (year of last certificate), by response and selected reported characteristics: 2013 .....	B-17
Table B.18.	Percentage of NATES respondents reporting specified responses to Q28a/b (certificate for past or future job), by response and selected reported characteristics: 2013.....	B-18
Table B.19.	Percentage of NATES respondents reporting specified responses to Q30 (year of most recent apprenticeship), by response and selected reported characteristics: 2013 ....	B-19
Table B.20.	Percentage of NATES respondents reporting specified responses to Q36 (classes to earn credential), by response and selected reported characteristics: 2013 .....	B-20
Table B.21.	Percentage of NATES respondents reporting specified responses to Q37 (credential earned), by response and selected reported characteristics: 2013 .....	B-21
Table B.22.	Percentage of NATES respondents reporting specified responses to Q50 (training provided by employer), by response and selected reported characteristics: 2013 .....	B-22



Table B.23.	Percentage of NATES respondents reporting specified responses to Q78 (how well English spoken), by response and selected reported characteristics .....	B-23
Table C.1.	Selected tract-level estimates, by number of missing auxiliary variables: 2013.....	C-2
Table C.2.	Odds ratios for probability of having 2 or more missing auxiliary variables, by selected tract-level estimate: 2013 .....	C-3
Table D.1.	Standard errors for table 2.1: Percentage distribution of NATES respondents and eligible sample and estimated bias, by weighting type and selected auxiliary variables: 2013 .....	D-1
Table D.2.	Standard errors for table 2.2: Percentage distribution of NATES plus NRFU respondents and NATES-only respondents and estimated bias, by weighting type and selected reported characteristics: 2013 .....	D-4
Table D.3.	Standard errors for table 2.3: Percentage distribution of NATES respondents, by weighting type and selected reported characteristics: 2013 .....	D-5
Table E.1.	Timeline of NATES and NRFU data collection activities: 2013 .....	E-1
Table E.2.	Number of contact attempts required for NRFU respondent households: 2013 .....	E-6
Table E.3.	Components of NATES household-level selection weights: 2014 .....	E-7
Table E.4.	Variables used in NATES household-level CHAID analysis: 2013 .....	E-9
Table E.5.	NATES household-level nonresponse adjustment cells: 2013.....	E-11
Table E.6.	Variables used in NATES person-level CHAID analysis: 2013 .....	E-14
Table E.7.	NATES person-level nonresponse adjustment cells: 2013 .....	E-16
Table E.8.	NATES household-level outcome codes, definition of each outcome code, and number of sampled households with each outcome code: 2013 .....	E-21
Table E.9.	NATES person-level outcomes: 2013.....	E-22
Table E.10.	NRFU outcome codes, definition of each outcome code, and number of sampled households with each outcome code: 2013 .....	E-23

*This page is intentionally left blank.*

# Chapter 1: Introduction

This report presents an analysis of unit and item nonresponse bias in the National Adult Training and Education Survey (NATES), a 2013 pilot study sponsored by the National Center for Education Statistics (NCES). The purpose of the NATES pilot study was to determine whether high-quality data on adults' education, training, and work-related credentials could be obtained using a mailed household survey. NCES was particularly interested in an evaluation of the extent of nonresponse bias in key NATES estimates based on the mailed survey. The NATES data collection included a methodological study to compare key estimates between households that did and did not respond to the mail questionnaire. The NATES Nonresponse Follow-up Study (NRFU) questionnaire was administered in person by trained field interviewers to a subsample of households that did not respond to the mailed questionnaire. The responses to this short survey with twelve key questions permit the measurement of unit nonresponse bias in a more direct fashion than is typically possible in federal household surveys. This report uses data from both the mailed NATES and the in-person NRFU.<sup>3</sup>

The remainder of this chapter proceeds as follows. Section 1.1 provides additional background information on the larger research effort of which NATES was a part. Section 1.2 describes the topical focus of the NATES and NRFU questionnaires. Section 1.3 introduces key methodological characteristics of the study and reports response rates for both the NATES and the NRFU. Section 1.4 introduces the concepts of unit and item nonresponse bias and methods that can be used to measure both, with an emphasis on the unique aspects of the methods used in this report. Finally, section 1.5 identifies the key research questions addressed in chapters 2–4 of this report.

The research effort described in this report was undertaken for questionnaire and procedural development purposes only. The information collected and published from this effort should not be used to generate or cite population estimates because the sample design and data collection procedures were not intended for that purpose, but rather to measure the potential for nonresponse bias in a household mail survey and to support the evaluation of questionnaire items.

## 1.1 Need for Data

NATES was the second of three pilot studies aimed at developing improved measures of the attainment of non-degree, work-related educational credentials and training—including certifications/licenses, educational certificates, and work experience programs—among U.S. adults. The impetus for this research effort was the recognition that existing federal data collection systems did not permit accurate and comprehensive measurement of economically valuable educational and training credentials other than 2-year, 4-year, and graduate degrees. The NCES Integrated Postsecondary Education Data System (IPEDS) collects data on educational certificates, but only those provided by postsecondary institutions that participate in federal student financial aid programs. Some certificates are awarded by educational

---

<sup>3</sup> Although the results shown in this report use data only from NATES and the NRFU, a number of other surveys are referenced at various points throughout the text. Full citation information for all surveys referenced in the text, including links to public-use datasets if available, is provided in the “References” section.

institutions that do not participate in these programs, and other types of industry-recognized credentials are typically awarded by noneducational institutions, such as trade associations, governments, and employers themselves. Given the sheer range and number of institutions that provide such credentials, an institution-based data collection would be infeasible. Interviewing adults through a household-based survey is a more efficient means of filling the gap in federal statistics on the attainment of non-degree educational and training credentials.

For this reason, the Interagency Working Group on Expanded Measures of Enrollment and Attainment (GEMEnA)<sup>4</sup> was created in 2009 to coordinate the development and validation of new survey items intended to measure such credentials. GEMEnA's portfolio includes supporting NCES in the development and piloting of a new household survey on education, training, and work-related credentials; the ultimate goal was to integrate this survey into the National Household Education Surveys Program (NHES), a repeated cross-sectional data collection sponsored by NCES. The first pilot survey sponsored by NCES as a part of this effort, the Adult Training and Education Survey (ATES), was fielded as a telephone survey in 2010. The ATES results informed further developmental work on key survey items, which led to the fielding of NATES in 2013.<sup>5</sup> The third pilot survey in the series was the NHES Feasibility Study, conducted in 2014 to test the integration of ATES into the NHES data collection methodology. Information on GEMEnA as well as questionnaires and reports from these pilot surveys can be found at [nces.ed.gov/surveys/gemena](http://nces.ed.gov/surveys/gemena).

The main purpose of NATES was to evaluate the feasibility of using a mailed survey to collect detailed data on education, training, and credentials from U.S. adults. Due to falling response rates for telephone surveys, NCES has shifted the NHES data collection program from telephone to mailed surveys. Furthermore, while most NHES administrations between 1991 and 2005 had incorporated an adult-focused survey, the adult component of the 2007 NHES (the last phone-based administration) was canceled due to low response rates. For these reasons, feasibility testing was necessary to determine whether a new adult-focused survey could eventually be reincorporated into the new mail-based NHES.

## 1.2 Topical Focus of the NATES Questionnaires

The target population for NATES was noninstitutionalized adults ages 16–65 who were no longer enrolled in grade 12 or below. The mailed NATES questionnaire included core substantive sections focused on each of the following topics: high school diplomas and college degrees, certifications/licenses, educational certificates, apprenticeships, and other work-related training. Within each section, the first item asked whether respondents possessed the associated credential or had participated in the type of training; subsequent items then requested detailed information about the credential or training, such as the field of study, the awarding institution, and its relevance to the respondent's current job. The questionnaire also included a section that asked whether respondents were currently enrolled in college-

---

<sup>4</sup> GEMEnA is comprised of staff from the following agencies: Census Bureau (U.S. Department of Commerce), Bureau of Labor Statistics (U.S. Department of Labor), Council of Economic Advisers (Executive Office of the President), National Center for Education Statistics (U.S. Department of Education), National Center for Science and Engineering Statistics (National Science Foundation), Office of Statistical and Science Policy (Office of Management and Budget), and Office of the Under Secretary (U.S. Department of Education).

<sup>5</sup> Only data from the NATES pilot (including the Nonresponse Follow-up Study incorporated into NATES) were used in the preparation of this report. For the results of the ATES:2010 pilot, see Bielick, Cronen, Stone, Montaquila, and Roth (2013).

level classes and then requested similar information about any reported classes. In this report, the term “key survey items” is used to identify the first item in each of the core substantive sections:

- Q1: “What is the highest degree or level of school you have completed?”
- Q4: “Do you have a professional certification or a state or industry license?”
- Q20: “Have you ever earned [an] educational certificate?”
- Q29: “Have you ever participated in [an] apprenticeship program?”
- Q35: “Are you currently taking classes from a vocational or trade school, community or technical college, or other college or university?”
- Q46: “Other than apprenticeships and college classes you may have described earlier, in the past 12 months, have you completed any other courses, training, or formal instruction, either at work or outside of work?”

In addition to these core substantive sections, information was requested on respondents’ employment status and demographic characteristics. A full copy of the NATES questionnaire can be found in appendix F of this report.

The in-person NRFU questionnaire was a substantially shortened version of the mailed NATES questionnaire. It included the key survey items that asked whether respondents possessed a certification, license, or certificate and whether respondents were currently enrolled in college-level classes, but it did not request detailed information on the characteristics of these credentials or classes. The NRFU questionnaire also included the NATES item on the attainment of high school diplomas or college degrees and several other key demographic items. It included no items related to apprenticeships or other work-related training or to respondents’ employment status. Appendix G provides a copy of the NRFU questionnaire.

### **1.3 NATES Sampling Frame, Data Collection, Weighting, and Response Rates**

Detailed documentation of the NATES sampling, data collection, weighting, and response rate calculation procedures is provided in appendix E. This section provides a basic introduction to the NATES methodology and reports response rates for both the main mailed NATES data collection phase and the in-person NRFU data collection phase.

#### ***Sampling frame***

The original sampling frame for NATES was the address-based sampling (ABS) frame maintained by Marketing Systems Group (MSG), which is based on the U.S. Postal Service’s (USPS) Computerized Delivery Sequence (CDS) file. The NATES sample was drawn from a set of 48,000 addresses that had originally been sampled for the 2012 National Household Education Survey (NHES:2012), but that had been randomly selected to be held in reserve and were thus unused.

In addition to a mailing address for each household, the NATES frame included a number of variables that provided some information on all or most sampled addresses prior to data collection. These included address-level characteristics provided by the USPS, such as the route type and vacancy status. They also included some information about the demographic characteristics of the household residing at the address (e.g., household income and the age and educational attainment of the “head of household”), which were

linked to the sampling frame by the vendor from several commercial databases. Throughout this report, the frame and commercial variables that were available for sampled cases *prior to data collection* are referred to jointly as “auxiliary variables.” In this way, the data that they provide for all or most sampled cases (regardless of the cases’ response status to NATES or the NRFU) are distinguished from the data collected only from survey respondents using the NATES or NRFU instrument.

Because auxiliary variables are available for both respondents and nonrespondents, they can (as discussed below) be used to adjust survey estimates for nonresponse and/or to analyze the risk of unit nonresponse bias. However, their usefulness is limited by several known quality problems (Harter et al. 2016). First, many auxiliary variables—particularly the demographic variables appended from commercial databases—show high missing rates. In the NATES sample, some of the variables used for nonresponse adjustment showed missing rates in excess of 40 percent (as shown in table E.4 of appendix E). Second, for addresses for which they are available, these variables are known to be subject to substantial measurement error. Using the same sampling frame as NATES, Roth, Han, and Montaquila (2013) compared the data provided in auxiliary variables to self-reported survey data and found numerous mismatches. For example, non-missing values of the commercial educational attainment variable matched self-reported educational attainment for 37 percent of survey respondents, and non-missing values of the income variable matched self-reported household income for 54 percent of survey respondents. The auxiliary variable for home tenure (own or rent) and a flag for Hispanic ethnicity were found to be more accurate, but still did not match self-reported data for 13 and 8 percent of respondents, respectively. DiSogra, Dennis, and Fahimi (2010) and Pasek et al. (2014) obtained similar results, finding particular inaccuracies (relative to self-reported survey data) in commercial auxiliary variables for income, household size, and education. Amaya, LeClere, Fiorio, and English (2014) found some inaccuracies (relative to field observations) in the address type variables provided by the USPS, particularly a flag for vacant addresses. A comprehensive review of other recent studies of the quality of commercial auxiliary data is provided by West, Wagner, Hubbard, and Gu (2015).

### **Data collection**

Data collection for NATES was conducted by the U.S. Census Bureau. In the first phase of data collection (the mailed phase), which began in January 2013, a stratified cluster sample of 10,000 households was selected from the NHES:2012 reserve sample. Each sampled household was mailed three copies of the NATES questionnaire.<sup>6</sup> The questionnaire instructions indicated that every member of the household between the ages of 16 to 65 who was no longer enrolled in high school should complete a questionnaire. Households with more than three eligible members were able to call a toll-free help line to request additional questionnaires, which were sent in the next scheduled follow-up mailing.

Nonresponding households received up to three follow-up mailings. The mailed phase of the NATES data collection ended on April 9, 2013. Eligible households that returned a questionnaire on or before April 9 were classified as NATES respondents. The approximately 3,610 sampled households that had neither

---

<sup>6</sup> For half of the sampled households, three separate booklets were provided; for the other half, all three copies of the questionnaire were contained in a single booklet. These two approaches were used as part of a separate randomized experiment that is not discussed in this report.

returned a questionnaire nor been classified as out of scope after April 9 were classified as NATES nonrespondents, and therefore eligible for the NRFU.

In the second phase of data collection, referred to as the Nonresponse Follow-up Study (NRFU), approximately 1,690 nonresponding households were selected for in-person interviews.<sup>7</sup> As discussed in detail in appendix E, nonrespondents were subsampled for the NRFU at rates that varied depending on the size of the Primary Sampling Unit (PSU) in which they were located. Field representatives were sent to the address provided in the sampling frame and asked to speak to any member of the household who met the NATES eligibility criteria. While the mailed phase had asked every eligible member of the household to complete and return a questionnaire, only one eligible person per household was interviewed for the NRFU. NATES nonrespondents who responded to this second phase were classified as NRFU respondents. Approximately 70 of the 1,690 households that were sampled for the NRFU subsequently returned a mailed questionnaire after the cutoff date; these households did not receive in-person interviews, but their responses to the mailed questionnaire items were included in the NRFU data file. These households were classified as NRFU respondents, *not* as NATES respondents.

## ***Weighting***

In survey designs that use probability sampling with sample members having a known probability of selection, each sample member is assigned a selection weight equal to the inverse of its probability of selection. The sum of the selection weights of eligible sample members is equal to the size of the target population. This is particularly important for stratified designs in which certain subpopulations are sampled at higher rates than others. In such cases, selection weights are needed to ensure that these populations are not overrepresented when calculating sample statistics.

Because NATES used a probability sample stratified by tract-level ethnicity and poverty rates, each sampled household was assigned a selection weight corresponding to its probability of selection. However, because not every sampled household returned a questionnaire, the sum of the selection weights of the NATES respondent households was lower than the eligible population count. For this reason, as is standard in NCES household surveys, the selection weights of NATES respondent households were adjusted upward in order to correct for unit nonresponse.

The NATES nonresponse adjustment method proceeded as follows. A data-mining algorithm called Chi-Squared Automatic Interaction Detection (CHAID) was used to allocate sampled households into subgroups—referred to as “adjustment cells”—defined by those characteristics that the algorithm found to be most strongly predictive of unit response status. Because an independent variable needed to be available for both respondents and nonrespondents in order for the CHAID algorithm to determine whether it was predictive of response status, the adjustment cells could be defined only by characteristics for which preexisting data were available for sampled households. For NATES, the variables used in the nonresponse adjustment process consisted of auxiliary data that were available in the sampling frame, or linked to the frame from commercial sources (refer to table E.4 of appendix E for a list of auxiliary

---

<sup>7</sup> This sample size was recommended by the U.S. Census Bureau based on a target of approximately 1,000 completed interviews and an assumed unweighted response rate of 60 percent.

variables used for the household-level nonresponse adjustment).<sup>8</sup> Each respondent household's selection weight was multiplied by the inverse of its adjustment cell's selection-weighted response rate in order to generate a household-level nonresponse-adjusted weight. The nonresponse-adjusted weights for nonrespondent and ineligible households were set equal to zero.

The household-level nonresponse-adjusted weights were then multiplied by an inflation factor in order to generate person-level selection weights for the individuals within NATES respondent households. The inflation factor was determined by the number of eligible persons in the household and the number of questionnaires that the household returned. The nonresponse adjustment procedure described above was then used to generate final person-level weights corrected for within-household nonresponse (refer to table E.6 of appendix E for a list of the auxiliary variables used for the person-level nonresponse adjustment).

The NATES nonresponse adjustment procedure reduces nonresponse bias to the extent that the auxiliary variables used to define the adjustment cells are also correlated with responses to the survey items. For example, the more strongly correlated the characteristics used to define the adjustment cells are with whether a person holds a professional certification or license, the more the NATES nonresponse adjustment procedure will reduce unit nonresponse bias in estimates of the population prevalence of professional certifications and licenses. Therefore, as discussed in detail in chapter 3, the effectiveness of this procedure could be hindered by high missing data rates for auxiliary variables and/or by weak underlying associations between the characteristics measured by the auxiliary variables and the characteristics measured by the survey instrument. Inaccuracies in the available auxiliary data, to the extent that they attenuate correlations that would otherwise exist between the auxiliary data and key outcomes, may also reduce the effectiveness of the nonresponse adjustments.

A separate weighting procedure was used for the NATES nonrespondent households that were sampled for the NRFU. These households' original selection weights were multiplied by the inverse of their probability of selection for the NRFU sample. Because of the high response rate to the in-person interviews (see below) and the lack of major differences between NRFU respondents and NRFU nonrespondents (see appendix A), nonresponse adjustments were not applied to the NRFU selection weights. Because only one person in each household was interviewed, and this person was not selected at random, no further person-level adjustment was made to the household-level selection weight for NRFU respondents.<sup>9</sup>

### ***NATES and NRFU response rates***

The mailed NATES questionnaire achieved a weighted household-level response rate of 65.0 percent, with 5,480 households responding. After accounting for within-household nonresponse, the final weighted person-level response rate was 62.5 percent, with a total of 7,540 persons responding. The

---

<sup>8</sup> In order to allow the assignment of households with missing auxiliary data to adjustment cells, "missing" was treated as its own category for each of the auxiliary variables used.

<sup>9</sup> Because the mailed NATES attempted to collect data from all eligible persons in each household, whereas the NRFU collected data only from the most readily available eligible person, some adjustments were made to the data to enable comparisons of NATES to NRFU data for the purpose of estimating nonresponse bias. These adjustments are described in detail in section E.5 of appendix E.



weighted response rate to the NRFU in-person interviews was 83.1 percent, with 1,180 households responding. The combined household-level response rate to NATES and the NRFU was 93.0 percent.

## 1.4 Methodological Introduction

### ***Unit nonresponse bias***

Broadly speaking, unit nonresponse bias is the deviation of a survey estimate from the true population value, attributable to the fact that not all sampled households or persons completed the survey questionnaire. A low response rate does not, by itself, imply the presence of unit nonresponse bias.

The magnitude of unit nonresponse bias in an estimate is determined not only by the rate of nonresponse, but also by differences between respondents and nonrespondents on the variable being measured. Specifically, nonresponse bias in a sample mean can be estimated as the product of the nonresponse rate and the difference between the mean among respondents and the mean among nonrespondents (Groves 1989):

$$Bias(y_r) = y_r - y_n = \frac{m}{n} (y_r - y_m)$$

where:

$y_r$  = the sample mean calculated using sample respondents

$y_n$  = the sample mean calculated using the full eligible sample

$y_m$  = the sample mean calculated using sample nonrespondents

$m$  = the number of nonrespondents in the sample

$n$  = the size of the full eligible sample

For example, despite the fact that the mailed NATES questionnaire did not achieve a 100 percent response rate, the NATES estimate of the population prevalence of certifications and licenses would be affected by unit nonresponse bias only if individuals who responded to NATES were either more or less likely to hold a certification or license than those who did not respond to NATES. A distinction can be made between ignorable nonresponse bias, which can be mitigated by weighting or otherwise conditioning an estimate on characteristics that are known for the entire sample or population; and nonignorable nonresponse bias, which remains even after the estimate is conditioned on available covariates (Rubin 1976; Little and Rubin 1987).

The *NCES Statistical Standards* (Seastrom 2012) require an analysis of unit nonresponse bias for any survey with a response rate below 85 percent. In most studies, the measurement of unit nonresponse bias is complicated by the fact that, by definition, survey data are unavailable for unit nonrespondents. This means that  $y_n$  and  $y_m$ , as defined above, cannot be calculated for any characteristic measured by the survey instrument. For these reasons, a number of indirect methods are often used to evaluate the risk of unit nonresponse bias.

One standard method of measuring unit nonresponse bias is to compare the survey respondents to the eligible sample in terms of auxiliary variables available in or linked to the sampling frame. This method is typically used when no additional information is available on the characteristics of nonrespondents. This method typically concludes that a risk of unit nonresponse bias exists if significant demographic differences exist between respondents and nonrespondents—in other words, if members of certain

subgroups are underrepresented in the respondent pool relative to their representation in the sample. The limitation of this method is that it allows the direct measurement of bias only in the characteristics available prior to data collection. Because these characteristics can be used to create nonresponse-adjusted weights (e.g., using the procedure described in section 1.3), bias in these characteristics, by definition, corresponds to ignorable nonresponse bias. In order to conclude on the basis of this method that unit nonresponse bias exists in the survey estimates, it is necessary to assume that individuals who differ in terms of the auxiliary variables also differ in terms of the characteristics measured by the questionnaire (i.e., that the auxiliary variables are correlated with the outcome measures).

A more thorough analysis of nonresponse bias is possible for NATES, because data for key survey items *are* available for NATES nonrespondents—specifically, the data from the in-person follow-up interviews. The selection-weighted respondents to the NATES questionnaire are, by definition, representative of the population of NATES respondents. The NRFU sample was selected and weighted to be representative of the population of NATES nonrespondents—that is, the population of which the selection-weighted nonrespondents to NATES are representative. Thus, a selection-weighted proportion generated using combined data from NATES and NRFU respondents can be treated, roughly speaking, as an estimate of the “true” population proportion, unaffected by nonresponse bias.<sup>10</sup> Nonresponse bias in the NATES proportion can therefore be estimated as the difference between the proportion generated using only NATES respondents and the same proportion generated using the combined NATES and NRFU respondents. This method of mitigating nonresponse bias through the use of more intensive data collection methods on a subsample of initial nonrespondents is well established in the survey research literature (cf. Hansen and Hurwitz 1946, Deming 1953, Tersine and Starsinic 2003, Harter et al. 2007).

Consequently, while chapter 2 of this report does present results for the indirect method described above, the bulk of the discussion focuses on the bias estimates calculated using the NRFU data.

### ***Item nonresponse bias***

Item nonresponse bias is bias in a particular survey estimate attributable to the fact that not all unit respondents respond to every item in the questionnaire. As is the case with unit nonresponse bias, item nonresponse bias is a function of the proportion of unit respondents who did not respond to the item and of differences between item respondents and nonrespondents in terms of the characteristic being measured by the item.

In NATES, weighted item missing rates for the key survey items were all below 3 percent; thus, the risk of item nonresponse bias was considered to be negligible for these items. However, as discussed in detail in chapter 4, missing rates were substantially higher for many of the items related to the characteristics of reported credentials. Because these items were skip-dependent—that is, respondents were instructed to either answer or skip them based on their response to a previous item—their high missing rates raise the possibility that some respondents may have had difficulty following the skip patterns in the NATES questionnaire. Complex skip patterns are of particular concern for mailed surveys, because respondents

---

<sup>10</sup> In practice, because the NRFU follow-up survey did not achieve a 100 percent response rate, the base-weighted sample of NRFU respondents may not be representative of the entire population of NATES nonrespondents; consequently, the measures of nonresponse bias calculated using this method may still be somewhat imprecise. However, an analysis reported in appendix A suggests that the risk of unit nonresponse bias in the NRFU itself is relatively low.

are required to navigate the survey themselves rather than with the assistance of an interviewer. For these reasons, the NATES item nonresponse bias analysis focuses on a selection of skip-dependent items.

For NATES, no follow-up was conducted with item nonrespondents, nor were the sampling weights adjusted for item nonresponse. Consequently, only indirect methods of measuring the risk of item nonresponse bias are feasible. For example, “extreme” values (typically the 5th and 95th percentiles of the response distribution among item respondents) can be imputed for item nonrespondents in order to analyze the resulting change in estimates and thereby establish lower and upper bounds on the extent of possible nonresponse bias for low-response items (cf. Roth, Montaquila, and Chapman 2006). Alternatively, for each low-response item, item respondents can be compared to item nonrespondents in terms of self-reported demographic characteristics (cf. Van de Kerckhove, Krenzke, and Mohadjer 2009). In this report, the primary interest is in determining whether item nonresponse led to certain subgroups being over- or underrepresented among item respondents and whether this, in turn, may have led to item nonresponse bias. For this reason, chapter 4 of this report uses a modified version of the latter method.

In particular, for selected low-response items, item respondents are compared to item nonrespondents in terms of age, educational attainment, employment status, and whether a language other than English is spoken in the household. Additionally, among item respondents, the association between these characteristics and the actual response to the item is evaluated. The combination of these analyses cannot determine for certain whether item nonresponse bias actually exists. However, it can determine whether item respondents and nonrespondents differ from one another in ways that could reasonably be expected to change the distribution of responses to the item (relative to the distribution that would have resulted had all participants answered the item). If a particular characteristic is predictive both of item response status and of the actual response to that item, it is suggestive of a risk of nonresponse bias for the item.

## 1.5 Key Research Questions and Structure of Report

This section provides a chapter-by-chapter summary of the seven research questions addressed in the report.

Chapter 2 addresses three research questions related to unit nonresponse bias in the NATES estimates:

- **Question 1: Measuring unit nonresponse bias before nonresponse adjustment.** To what extent are key survey estimates derived from selection-weighted NATES respondents affected by unit nonresponse bias?
- **Question 2: Assessing the efficacy of nonresponse adjustment.** Is unit nonresponse bias substantially reduced when the sampling weights for NATES respondents are adjusted for nonresponse?
- **Question 3: Comparing alternative methods of unit nonresponse bias analysis.** Do methods that are available for nonresponse bias analysis in the absence of an NRFU provide an accurate indication of the risk of bias and of the effectiveness of nonresponse adjustment?

Following up on the findings reported in chapter 2, chapter 3 evaluates two potential methods of reducing unit nonresponse bias in future administrations of the Adult Training and Education Survey (ATES)<sup>11</sup> and other household surveys:

- **Question 4: Improving weighting adjustments for nonresponse.** Is there reason to expect that the effectiveness of the weighting adjustments at reducing nonresponse bias could be improved in future administrations if more complete data were obtained for the demographic characteristics linked to the sampling frame?
- **Question 5: Increasing response rates among key subgroups.** Is there reason to expect that nonresponse bias in future administrations would be reduced if certain subgroups of sampled households were chosen to receive larger incentives, more intensive nonresponse follow-up, or other targeted efforts to increase response rates?

Chapter 4 analyzes item nonresponse bias in a selection of items that immediately followed skip directives and that showed low response rates (below 85 percent). In particular, the chapter addresses two research questions:

- **Question 6: Identifying differences between item respondents and nonrespondents.** Were any demographic characteristics consistently associated with nonresponse to skip-dependent items?
- **Question 7: Assessing the risk of item nonresponse bias.** For each of the items evaluated, is there reason to expect that differences between item respondents and nonrespondents led to changes in the distribution of reported responses to the item?

The report concludes with chapter 5, which summarizes the implication of these analyses for future administrations of ATES and other mailed household surveys, and offers recommendations for the amelioration of unit and item nonresponse bias in such surveys.

The scope of this report is limited to the analysis of unit and item nonresponse bias. Other types of nonsampling bias, including measurement error and coverage bias, are not considered. Other potential limitations of this study are discussed in section 5.2 of chapter 5.

---

<sup>11</sup> The first full-scale administration of the new adult education survey took place as part of the 2016 National Household Education Survey. NCES has reverted to the use of ATES as the official title of the survey, removing the word “National” to avoid redundancy. Consequently, in this report, all post-NATES administrations of adult-focused surveys based on the developmental work of which NATES was a part are referred to as ATES administrations.

## Chapter 2: Measuring Unit Nonresponse Bias

Sections 2.1 through 2.3 use multiple methods to address the report’s first three research questions, a full list of which is provided in section 1.5 of chapter 1.

In particular, section 2.1 evaluates differences between the respondent pool and the entire eligible sample in terms of demographic and other characteristics available in or linked to the sampling frame. Section 2.2 uses data from the NRFU follow-up sample to estimate unit nonresponse bias in items that were included in both the NATES and the NRFU questionnaires. Finally, for several key NATES items not included in the NRFU questionnaire, section 2.3 compares selection-weighted estimates to nonresponse-adjusted estimates in order to provide additional insight into the effectiveness of the nonresponse adjustment process. In these sections, unless otherwise stated, the estimated bias is considered to be statistically significant if the  $p$  value of its Student’s  $t$  test is less than .05, in accordance with the *NCES Statistical Standards* (Seastrom 2012). All  $t$  tests presented in this chapter account for the overlap between the full sample and the respondents by using a jackknife replication method to directly estimate the standard error of the bias. Additional detail about the estimation of standard errors is provided in section E.3 of appendix E. The  $t$  tests are treated as being independent of each other, and no adjustment for multiple comparisons is made. Because of the large sample size, and the likely correlation between the full-sample and respondents-only estimates, relatively small differences could be statistically significant; for this reason, a substantive importance threshold is also used. The estimated bias is considered to be substantively important if its absolute value exceeds one percentage point, the same standard used in the NHES:2012 nonresponse bias analysis (McPhee, Bielick, Masterton, Flores, Parmer, Amchin, Stern, and McGowan 2015).

### 2.1 Evaluation of Bias Using Auxiliary Variables

As discussed in chapter 1, a standard method of evaluating nonresponse bias in the absence of a nonresponse follow-up study is to compare respondents to the eligible sample as a whole in terms of the percentage distributions of auxiliary variables available in or linked to the sampling frame. This method cannot directly measure unit nonresponse bias in the NATES estimates, but it does provide insight into which demographic subgroups were more likely to respond to the mailed NATES questionnaire.

Table 2.1 compares the percentage distribution of the NATES eligible sample to the percentage distribution of NATES respondents in terms of 15 auxiliary variables, for a total of 68 categories. Bias in each proportion is estimated as the difference between the proportion estimated for the respondents-only group ( $p_r$ ) and the proportion estimated for the entire eligible sample ( $p_e$ ):

$$B(p_r) = p_r - p_e$$

The eligible sample proportions (column 1) are estimated using selection weights. The respondent proportions are estimated in column 2 using selection weights to create the measure of bias in each proportion prior to nonresponse adjustment (column 3). The respondent proportions are then re-estimated

using nonresponse-adjusted weights (column 6), after which the measure of bias in each proportion is recalculated (column 7). This allows an evaluation of the extent to which the nonresponse adjustments reduce differences between respondents and the eligible sample in terms of the distribution of auxiliary variables.

The statistical significance of the bias in each proportion is evaluated using a Student's  $t$  test in which the  $t$  statistic (columns 5 and 10) is defined as the estimated bias divided by the standard error of the bias. Since NATES had a complex sample design, the standard error of the bias is calculated using the jackknife method with 70 replicates (see appendix D for standard error tables). In addition, the estimated bias in each proportion is transformed into the percent relative bias (columns 4 and 8), defined as the ratio of the estimated bias to the estimate of the eligible sample proportion. The relative bias is independent of the distribution of particular variables and, therefore, allows the comparison of bias across all variables. Finally, a statistic called the percent relative difference (column 9), defined as the percentage change in the absolute value of the estimated bias, provides a common measure of the change in bias attributable to the use of nonresponse-adjusted weights.

When the percentage distributions for respondent households are estimated using selection weights, all variables of interest (with the exception of the NATES form type) show statistically significant and substantively important bias in the estimated proportions for at least some subcategories. Relative to their share of the eligible sample, the following types of households are significantly underrepresented in the respondent pool:

- households located in high-minority strata;
- households for which the head of the household is Black or Hispanic;
- households with a low annual income;
- households that are located in high rises or receive mail at a P.O. box;
- households located in high-poverty tracts;
- households for which no phone number is available;
- households that include only one adult;
- households that rent their homes;
- households located in the South;
- households located in multi-unit dwellings; and
- households for which the head of the household is single.

Additionally, for every auxiliary variable for which some households are missing data, respondent households are significantly less likely (by at least 1.3 percentage points) to be missing data for that variable. This means that households with missing data for one or more of these variables were less likely to return the mailed NATES questionnaire. As discussed by Harter et al. (2016), and in appendix C of this report, missing data in appended auxiliary variables tends to reflect other characteristics that are likely to be associated with low response rates—for example, addresses in multi-unit structures, or in

neighborhoods with low incomes and/or highly mobile populations, are less likely to be successfully matched to commercial databases. Thus, the underrepresentation of households with missing auxiliary data among NATES respondents is consistent with these findings.<sup>12</sup>

---

<sup>12</sup> While the estimates of bias in the non-missing categories could, in principle, be distorted by the inclusion of “missing” as a category in the percentage distributions, a sensitivity analysis (not shown in tables) confirmed that the conclusions discussed in this section remain robust when cases with missing data are excluded from the percentage distributions for all variables.

**Table 2.1. Percentage distribution of NATES respondents and eligible sample, estimated bias, and percent relative bias, by weighting type and selected auxiliary variables: 2013**

Selected auxiliary variable	With selection weights					With nonresponse-adjusted weights				
	Percent of eligible sample 1	Percent of respondents 2	Estimated bias 3	Percent relative bias 4	t test of bias 5	Percent of respondents 6	Estimated bias 7	Percent relative bias 8	Percent relative difference in bias 9	t test of bias 10
Race/ethnicity stratum										
Black	14.8	12.9	-1.9	-12.8	-6.0*	13.6	-1.1	-7.6	-40.2	-3.4*
Hispanic	9.6	7.5	-2.1	-21.4	-9.7*	8.2	-1.4	-14.3	-33.5	-5.7*
Other	75.7	79.6	3.9	5.2	10.8*	78.2	2.5	3.3	-36.7	6.2*
Ethnicity of the head of household										
Missing	36.7	32.4	-4.3	-11.8	-8.7*	35.0	-1.7	-4.7	-60.4	-3.4*
White	44.4	49.6	5.2	11.8	10.1*	46.9	2.5	5.7	-51.8	4.6*
Black	6.8	6.3	-0.5	-7.4	-2.3*	6.3	-0.5	-7.3	-1.7	-2.1*
Hispanic	6.4	5.5	-0.9	-13.7	-3.6*	5.7	-0.7	-10.7	-21.4	-2.5*
Asian/Pacific Islander	2.6	2.8	0.2	7.1	1.4	2.8	0.2	7.0	-1.4	1.3
Other	3.1	3.4	0.3	9.9	2.0*	3.3	0.2	6.3	-36.0	1.3
Household income										
Missing	15.6	11.2	-4.4	-28.2	-10.0*	13.5	-2.2	-13.9	-50.8	-4.6*
\$0-\$10,000	2.7	2.0	-0.7	-25.7	-3.1*	2.4	-0.3	-11.0	-57.1	-1.2
\$10,001-\$20,000	5.7	4.8	-0.9	-15.5	-3.0*	5.3	-0.4	-7.3	-53.2	-1.3
\$20,001-\$30,000	8.3	7.9	-0.4	-5.1	-1.2	7.9	-0.4	-5.1	-0.2	-1.4
\$30,001-\$40,000	9.2	8.6	-0.5	-6.0	-1.9	9.1	0.0	-0.2	-96.3	-0.1
\$40,001-\$50,000	9.5	10.0	0.5	4.8	1.1	9.9	0.4	3.8	-20.9	0.9
\$50,001-\$60,000	8.9	9.1	0.2	1.9	0.5	9.2	0.3	3.0	62.0	0.8
\$60,001-\$75,000	10.8	11.8	1.0	9.5	2.0*	11.2	0.4	4.0	-58.1	0.9
\$75,001-\$100,000	12.8	14.5	1.7	13.5	5.2*	13.6	0.9	6.7	-50.3	2.6*
\$100,001-\$150,000	11.3	14.0	2.6	23.1	8.1*	12.5	1.2	10.5	-54.8	4.0*
\$150,001+	5.2	6.2	1.0	18.2	5.1*	5.4	0.2	4.1	-77.6	1.2

See notes at end of table.



**Table 2.1. Percentage distribution of NATES respondents and eligible sample, estimated bias, and percent relative bias, by weighting type and selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	With selection weights					With nonresponse-adjusted weights				
	Percent of eligible sample	Percent of respondents	Estimated bias	Percent relative bias	t test of bias	Percent of respondents	Estimated bias	Percent relative bias	Percent relative difference in bias	t test of bias
	1	2	3	4	5	6	7	8	9	10
Route type										
High rise	19.6	14.8	-4.8	-24.6	-11.1*	19.6	0.0	0.0	-99.8	0.0
Street	71.7	77.9	6.2	8.6	12.9*	74.0	2.3	3.3	-61.9	4.7*
Rural route	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
P.O. box	8.3	6.9	-1.3	-16.0	-3.5*	6.1	-2.2	-26.7	67.2	-6.3*
Education of the head of household										
Missing	40.4	35.7	-4.7	-11.6	-8.1*	38.7	-1.6	-4.0	-65.3	-2.7*
Less than high school diploma	10.2	9.8	-0.4	-4.2	-1.1	9.7	-0.5	-4.5	6.8	-1.2
High school diploma	16.8	18.3	1.5	8.7	3.3*	17.3	0.5	2.9	-66.9	1.1
Some college	15.6	16.1	0.4	2.9	1.2	15.9	0.3	1.8	-35.7	0.8
Bachelor's degree	10.6	12.4	1.8	17.0	6.1*	11.3	0.7	6.9	-59.1	2.5*
Graduate degree	6.4	7.8	1.4	21.6	4.7*	7.0	0.6	8.9	-58.9	2.0*
Age of the head of household										
Missing	38.6	31.1	-7.5	-19.4	-14.4*	36.3	-2.3	-6.0	-69.3	-4.0*
18–24	1.3	1.1	-0.2	-17.1	-1.8	1.2	-0.1	-5.8	-66.3	-0.6
25–34	6.0	5.7	-0.3	-5.0	-1.2	5.7	-0.3	-4.9	-3.5	-1.2
35–44	11.0	11.4	0.4	3.5	0.9	10.6	-0.4	-3.5	-1.2	-1.0
45–54	14.5	15.3	0.8	5.7	2.3*	14.9	0.4	2.9	-49.5	1.2
55–65	15.0	17.6	2.5	16.9	6.8*	16.7	1.6	10.9	-35.6	4.4*
Over 65	13.6	17.8	4.2	31.3	10.6*	14.6	1.0	7.3	-76.8	3.4*
Census tract poverty rate										
20 percent or higher	17.6	14.4	-3.2	-17.9	-7.5*	15.7	-1.9	-10.7	-40.4	-4.1*
Below 20 percent or missing	82.4	85.6	3.2	3.8	7.5*	84.3	1.9	2.3	-40.4	4.1*

See notes at end of table.

**Table 2.1. Percentage distribution of NATES respondents and eligible sample, estimated bias, and percent relative bias, by weighting type and selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	With selection weights					With nonresponse-adjusted weights				
	Percent of eligible sample	Percent of respondents	Estimated bias	Percent relative bias	t test of bias	Percent of respondents	Estimated bias	Percent relative bias	Percent relative difference in bias	t test of bias
	1	2	3	4	5	6	7	8	9	10
Phone number available in sampling frame										
Yes	41.5	48.2	6.7	16.1	12.6*	44.1	2.5	6.1	-62.0	4.8*
No	58.5	51.8	-6.7	-11.4	-12.6*	55.9	-2.5	-4.3	-62.0	-4.8*
NATES form type										
Individual	49.8	50.0	0.2	0.3	0.2	49.9	0.1	0.2	-47.9	0.1
Booklet	50.2	50.0	-0.2	-0.3	-0.2	50.1	-0.1	-0.2	-47.9	-0.1
Number of adults in household										
Missing	15.9	11.3	-4.5	-28.7	-10.1*	13.6	-2.3	-14.5	-49.6	-4.8*
1	34.3	31.6	-2.7	-7.8	-4.5*	34.1	-0.2	-0.7	-90.8	-0.4
2	27.6	31.8	4.2	15.4	8.1*	29.0	1.4	5.1	-66.9	2.8*
3	13.4	15.1	1.8	13.2	4.4*	14.2	0.8	6.3	-52.3	2.4*
4	5.7	6.5	0.8	13.5	2.7*	5.8	0.1	1.6	-88.3	0.4
5	2.1	2.5	0.3	16.0	2.1*	2.3	0.2	9.4	-41.4	1.2
6	0.8	0.9	0.1	11.6	1.2	0.8	0.0	1.9	-83.9	0.2
7	0.2	0.3	0.1	38.9	2.6*	0.3	0.1	26.1	-32.8	2.0*
8	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
Home tenure										
Missing	20.8	15.5	-5.3	-25.7	-11.9*	18.1	-2.6	-12.7	-50.4	-5.3*
Own	61.2	70.8	9.6	15.7	17.5*	64.2	3.1	5.0	-68.2	5.1*
Rent	18.0	13.7	-4.3	-23.7	-11.4*	17.6	-0.4	-2.3	-90.3	-0.9

See notes at end of table.

**Table 2.1. Percentage distribution of NATES respondents and eligible sample, estimated bias, and percent relative bias, by weighting type and selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	With selection weights					With nonresponse-adjusted weights				
	Percent of eligible sample 1	Percent of respondents 2	Estimated bias 3	Percent relative bias 4	t test of bias 5	Percent of respondents 6	Estimated bias 7	Percent relative bias 8	Percent relative difference in bias 9	t test of bias 10
Census region										
Northeast	18.0	18.4	0.3	1.8	0.8	18.2	0.1	0.6	-67.7	0.3
South	38.7	37.1	-1.6	-4.1	-2.7*	36.8	-1.9	-5.0	20.3	-3.3*
Midwest	21.3	23.0	1.7	8.2	3.5*	22.7	1.5	6.9	-16.2	2.9*
West	22.0	21.5	-0.5	-2.2	-0.8	22.4	0.3	1.6	-29.7	0.6
Dwelling type										
Missing	8.3	6.9	-1.3	-16.0	-3.5*	6.1	-2.2	-26.7	67.2	-6.3*
Single-family unit	70.0	76.6	6.6	9.4	13.5*	72.5	2.5	3.5	-62.6	4.7*
Multi-unit	21.7	16.5	-5.3	-24.2	-11.7*	21.5	-0.3	-1.2	-95.1	-0.5
Gender of the head of household										
Missing	18.1	12.9	-5.2	-28.8	-11.3*	15.5	-2.6	-14.6	-49.5	-5.2*
Female	26.1	25.2	-0.9	-3.5	-1.8	26.2	0.1	0.5	-87.1	0.2
Male	55.8	61.9	6.1	11.0	10.3*	58.3	2.5	4.5	-59.0	4.2*
Marital status of the head of household										
Missing	31.2	26.7	-4.5	-14.4	-8.3*	28.7	-2.5	-8.1	-43.6	-4.5*
Married	47.1	53.7	6.6	14.0	11.8*	49.4	2.3	4.8	-65.9	4.0*
Single	21.7	19.6	-2.1	-9.7	-5.1*	22.0	0.3	1.3	-86.7	0.6

‡ Reporting standards not met. Either there are too few cases for a reliable estimate or the coefficient of variation is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of eligible sampled households or respondent households with the specified characteristic. "Estimated bias" represents the difference between the respondent proportion and the eligible sample proportion. "Percent relative bias" represents the estimated bias divided by the eligible sample proportion. "Percent relative difference in bias" represents the percentage change in the absolute value of the estimated bias when nonresponse-adjusted weights are used. Details may not sum to totals due to rounding. Standard errors of the eligible sample and respondent proportions, as well as the estimated bias, are shown in table D.1 of appendix D.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

The nonresponse adjustments do appear to reduce bias in auxiliary variables. When selection weights are used, 40 proportions show bias that is both statistically significant and substantively important; when nonresponse-adjusted weights are used, the number of proportions showing statistically significant and substantively important bias decreases to 28. For 9 of the 15 variables analyzed, the absolute value of the estimated bias decreases by over 50 percent for the majority of the variable's categories. This is consistent with expectations, because 7 of these 9 variables were used to define the household nonresponse adjustment cells during the NATES weighting process. In total, 36 categories (31 of which were used to define the nonresponse adjustment cells) show a decrease of over 50 percent in the absolute value of the estimated bias; 3 show an increase of over 50 percent.

However, for almost all of the variables of interest (with the exception of the NATES form type), statistically significant and substantively important bias remains in the estimated proportions for at least a few categories. Overall, when all 68 categories from all 15 variables in table 2.1 are considered, the nonresponse adjustments reduce the absolute value of the estimated bias by an average of 40.7 percent and reduce the median percent relative bias (in absolute value terms) from 12.3 percent to 5.0 percent.

In summary, there appear to be significant differences in the characteristics of the responding and nonresponding households, as measured by auxiliary variables. Many, but not all, of these differences are reduced or eliminated when the nonresponse-adjusted weights are applied. However, differences in auxiliary characteristics between responding and nonresponding households are of interest only to the extent that they lead to unit nonresponse bias in the actual survey variables. In order to directly measure bias in survey variables, it is necessary to use data from the NRFU follow-up sample. Furthermore, because many of the auxiliary variables were themselves used to define the nonresponse adjustment cells, the apparent reduction in these variables' bias from the nonresponse adjustment procedure may overestimate the reduction in bias in the survey variables. Data from the NRFU sample will therefore provide more accurate insight into the effectiveness of the nonresponse adjustment procedure at removing bias in survey variables.

## 2.2 Measurement of Bias Using the NRFU Follow-Up Sample

Table 2.2 reports measures of unit nonresponse bias for 26 proportions from nine key items that were included in both the NATES and the NRFU questionnaires. Because the NRFU allows bias to be measured for characteristics that are not available in the sampling frame or from extant population data, and that therefore cannot be used in weighting adjustments, it allows an evaluation of nonignorable bias.

For each item, the percentage distribution of responses to the item is estimated using combined data from respondents to the mailed NATES survey and the NRFU in-person interviews. These proportions are referred to as the "NATES plus NRFU proportions." Because the combined household-level response rate to NATES and the NRFU was approximately 93.0 percent,<sup>13</sup> the NATES plus NRFU proportions are assumed to be reasonable approximations of the true proportions for the eligible population. The NATES plus NRFU proportions are compared to proportions estimated using only respondents to the mailed NATES survey—which are referred to as the "NATES-only proportions." For each proportion, unit

---

<sup>13</sup> The household-level response rate is cited because the analysis in this section is conducted at the household level.

nonresponse bias is estimated as the difference between the NATES-only proportion ( $p_n$ ) and the NATES plus NRFU proportion ( $p_c$ ):

$$B(p_n) = p_n - p_c$$

The NATES plus NRFU proportions (column 1) are estimated using selection weights. The NATES-only proportions are first estimated using selection weights (column 2) in order to measure the extent of bias in each proportion prior to nonresponse adjustment (column 3). The NATES respondent proportions are then re-estimated using nonresponse-adjusted weights (column 6), after which the measure of bias in each proportion is recalculated (column 7). This allows an evaluation of the extent to which the nonresponse adjustments reduce bias in the proportions.

As in the previous section, the statistical significance of each bias estimate is evaluated using a  $t$  test (columns 5 and 10), and the percent relative difference (column 9) shows the change in the absolute value of the bias attributable to the nonresponse adjustments. Also as in the previous section, the estimated bias in each proportion is standardized to the percent relative bias (columns 4 and 8), defined here as the ratio of the estimated bias to the NATES plus NRFU proportion.

Although every eligible person in each sampled household was asked to complete the mailed NATES questionnaire, only one eligible person per household was interviewed for the NRFU follow-up questionnaire. The interviewed person was not selected randomly; rather, the first available eligible respondent in the household was interviewed. This implies that person-level estimates cannot be obtained from the NRFU, because the use of a weighting adjustment for within-household selection implies random or at least quasi-random selection. For this reason, the NRFU was treated for analytical purposes as a household-level dataset that provides data for the “first, most convenient” reporter in the household. For example, the proportion of NRFU respondents reporting a certification/license is interpreted not as the proportion of persons with a certification/license, but rather as the proportion of households for which the “first, most convenient” reporter has a certification/license. Consequently, to allow for comparability between NATES and NRFU proportions, it was necessary to identify the equivalent to the “first, most convenient” reporter in households that returned more than one mailed NATES questionnaire. In practice, for NATES respondent households with more than one respondent, data were retained for the individual who was deemed the most likely to have been the first to fill out a questionnaire. Thus, the analysis in table 2.2 should be interpreted as evaluating bias in household-level proportions that are based on the “first, most convenient” reporter in each NATES and NRFU respondent household. The weights used in the analysis are therefore household-level, not person-level, selection and nonresponse-adjusted weights. A more detailed description of the procedure used to identify the “first, most convenient” reporter in NATES-responding households, and of the rationale for this approach, is provided in section E.5 of appendix E.

As shown in table 2.2, the estimated bias in the certification/license and certificate items is not statistically significant. The estimated bias in the college class enrollment item is statistically significant; the NATES-only proportions underestimate enrollment in college classes by about 1.5 percentage points. The percent relative bias in the selection-weighted college enrollment rate is 15.9 percent.

Statistically significant and substantively important unit nonresponse bias is more common for demographic characteristics, particularly those with a large number of subcategories. The NATES-only

proportions overestimate the proportion with a bachelor's or graduate degree by about 2 percentage points. Estimates of bias for the age, race, origin, and income variables are larger: the proportion over the age of 55 is overestimated by about 7 percentage points; the White proportion by about 5 percentage points; the non-Hispanic proportion by about 4 percentage points; and the proportion in the highest household income category (over \$75,000) by about 5 percentage points. Note that the demographic characteristics shown in table 2.2 represent responses to survey items, rather than the auxiliary variables available for the full sample that were used to adjust the weights for nonresponse.

**Table 2.2. Percentage distribution of NATES plus NRFU respondents and NATES-only respondents, estimated bias, and percent relative bias, by weighting type and selected reported characteristics: 2013**

Selected reported characteristic	With selection weights					With nonresponse-adjusted weights				
	Percent of NATES plus NRFU respondents 1	Percent of NATES-only respondents 2	Estimated bias 3	Percent relative bias 4	t test of bias 5	Percent of NATES-only respondents 6	Estimated bias 7	Percent relative bias 8	Percent relative difference in bias 9	t test of bias 10
Has a professional certification or license										
Yes	30.7	31.9	1.2	3.9	1.3	31.4	0.7	2.3	-40.0	0.8
No	69.3	68.1	-1.2	-1.7	-1.3	68.6	-0.7	-1.0	-40.0	-0.8
Has a certificate										
Yes	16.2	15.3	-0.8	-5.2	-1.0	15.2	-1.0	-6.2	19.2	-1.2
No	83.8	84.7	0.8	1.0	1.0	84.8	1.0	1.2	19.2	1.2
Currently enrolled in college classes										
Yes	9.3	7.8	-1.5	-15.9	-2.3*	8.3	-0.9	-10.1	-36.3	-1.4
No	90.7	92.2	1.5	1.6	2.3*	91.7	0.9	1.0	-36.3	1.4
Highest educational attainment										
Less than high school diploma	8.3	7.1	-1.2	-14.3	-2.7*	7.5	-0.8	-9.6	-32.6	-1.8
High school diploma or equivalent	25.3	22.8	-2.5	-9.7	-3.4*	23.0	-2.3	-8.9	-8.2	-3.1*
Some college or associate's degree	33.5	33.0	-0.5	-1.5	-0.7	32.9	-0.6	-1.7	14.6	-0.8
Bachelor's degree	20.5	22.3	1.8	8.9	3.4*	22.0	1.5	7.5	-15.6	2.9*
Graduate or professional degree	12.5	14.8	2.3	18.5	4.5*	14.6	2.1	16.7	-9.8	4.1*
Sex										
Male	44.6	43.3	-1.3	-3.0	-1.7	43.2	-1.4	-3.2	7.0	-1.7
Female	55.4	56.7	1.3	2.4	1.7	56.8	1.4	2.6	7.0	1.7

See notes at end of table.

**Table 2.2. Percentage distribution of NATES plus NRFU respondents and NATES-only respondents, estimated bias, and percent relative bias, by weighting type and selected reported characteristics: 2013—Continued**

Selected reported characteristic	With selection weights					With nonresponse-adjusted weights				
	Percent of NATES plus NRFU respondents	Percent of NATES-only respondents	Estimated bias	Percent relative bias	t test of bias	Percent of NATES-only respondents	Estimated bias	Percent relative bias	Percent relative difference in bias	t test of bias
	1	2	3	4	5	6	7	8	9	10
Age										
16–24	8.4	5.4	-3.1	-36.2	-5.3*	6.0	-2.4	-28.3	-21.9	-3.9*
25–34	19.3	14.2	-5.0	-26.1	-5.6*	15.5	-3.7	-19.4	-25.8	-4.1*
35–44	20.1	19.5	-0.6	-3.1	-0.9	19.4	-0.7	-3.4	12.8	-1.1
45–54	25.3	26.9	1.6	6.2	1.9	26.0	0.7	2.7	-55.7	0.9
55–65	26.8	34.0	7.1	26.6	9.1*	33.0	6.1	22.8	-14.3	7.9*
Race										
White	76.4	81.6	5.1	6.7	6.7*	80.4	4.0	5.2	-21.9	5.0*
Black	14.5	10.0	-4.5	-30.9	-6.6*	10.8	-3.7	-25.6	-17.2	-5.2*
Other or multiple	9.1	8.4	-0.7	-7.2	-1.3	8.8	-0.3	-3.3	-54.1	-0.6
Origin										
Hispanic	13.4	9.3	-4.1	-30.5	-5.3*	10.1	-3.3	-24.5	-19.6	-4.2*
Not Hispanic	86.6	90.7	4.1	4.7	5.3*	89.9	3.3	3.8	-19.6	4.2*
Household income										
\$0–\$30,000	28.3	25.0	-3.3	-11.6	-4.3*	26.7	-1.6	-5.8	-50.4	-2.1*
\$30,001–\$75,000	38.6	37.2	-1.4	-3.6	-1.6	37.7	-0.9	-2.5	-31.9	-1.1
\$75,001+	33.1	37.8	4.7	14.2	5.5*	35.7	2.6	7.8	-44.9	3.0*

\* $p < .05$ .

NOTE: Percentages represent the proportion of NATES plus NRFU respondents or NATES-only respondents with the specified characteristic. "Estimated bias" represents the difference between the NATES-only proportion and the NATES plus NRFU proportion. "Percent relative bias" represents the estimated bias divided by the NATES plus NRFU proportion. "Percent relative difference in bias" represents the percentage change in the absolute value of the estimated bias when nonresponse-adjusted weights are used. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding. Standard errors of the NATES plus NRFU proportions and the NATES-only proportions, as well as the estimated bias, are shown in table D.2 of appendix D. SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.



When nonresponse-adjusted weights are used for the NATES-only proportions, the estimated downward bias in the college class enrollment item decreases to about 0.9 percentage points, which is no longer statistically significant or substantively important. For the demographic proportions that showed the most bias when selection weights were used, the nonresponse adjustments achieve only moderate reductions in bias. For educational attainment, bias in the bachelor's and graduate degree proportions decreases by about 16 and 10 percent, respectively. For age, bias in the over-55 proportion decreases by about 14 percent. For the race and origin variables, bias in the estimated proportions of Whites and non-Hispanics decreases by about 22 and 20 percent, respectively. The overestimate in the highest income category is reduced more substantially, by about 45 percent.

Altogether, the nonresponse adjustments reduce the absolute value of the estimated bias by an average of 19.9 percent (29.8 percent if the estimates for which the adjustments increased the bias are excluded) and reduce the median percent relative bias (in absolute value terms) from 7.0 percent to 5.5 percent. When nonresponse-adjusted weights are used, the number of NATES estimates with statistically significant and substantively important bias decreases from 15 to 12. If only the three key items of interest (certification/license attainment, certificate attainment, and college enrollment) are considered, the average reduction in the absolute value of the estimated bias is 19.1 percent, and the median percent relative bias (in absolute value terms) is reduced from 2.8 to 1.8 percent.

In summary, no statistically significant bias was found in the certification/license and certificate items, with or without the use of nonresponse-adjusted weights. The college enrollment item showed statistically significant bias when estimated with selection weights, but not when estimated with nonresponse-adjusted weights. However, the demographic proportions that showed the most bias prior to nonresponse adjustment generally showed the lowest percentage reductions in bias when nonresponse adjustments were applied. When the selection-weighted NATES-only proportions are compared to the nonresponse-adjusted NATES-only proportions, it is clear that the nonresponse adjustments led, in general, to very small changes in the NATES estimates.

These results indicate that the nonresponse adjustment procedure used for NATES was, by itself, insufficient to remove a substantial portion of the nonresponse bias in demographic proportions. That this is true in spite of the fact that auxiliary variables purporting to capture these same demographic characteristics (e.g., race/ethnicity, age, educational attainment, etc.) were used in nonresponse adjustment is likely a reflection of the high missing rates and possible measurement error in the auxiliary variables (discussed in section 1.3 of chapter 3). However, the demographic characteristics shown in table 2.2 all reflect characteristics for which high-quality external estimates for the U.S. population can be obtained from sources such as the decennial census and the American Community Survey (ACS). The bias that remains in these variables after nonresponse adjustment could therefore be removed (i.e. rendered ignorable) through the use of an additional poststratification or raking adjustment to the weights. Because the NATES study aimed specifically to evaluate the nonresponse adjustment procedure based on the auxiliary data, no poststratification or raking adjustment was used; however, a raking adjustment will be incorporated into the weighting procedures for future full-scale ATES administrations.

## 2.3 Evaluation of Nonresponse Adjustments for NATES-Only Items

For items that were included in the mailed NATES questionnaire but not in the NRFU follow-up questionnaire, it is not possible to directly measure nonresponse bias. For this reason, it is not possible to determine the extent to which the nonresponse adjustments were effective at reducing nonresponse bias in these items. However, under the assumption that an unknown amount of unit nonresponse bias exists for these items, some insight can be gained by comparing selection-weighted proportions to nonresponse-adjusted proportions. If the nonresponse adjustments lead to only slight changes in the estimated proportions, then they may not be particularly effective at reducing nonresponse bias in these proportions.

Table 2.3 reports selection-weighted and nonresponse-adjusted proportions for the attainment of the two major credentials of interest—apprenticeships and work-related training—that were included on the NATES questionnaire, but not included on the NRFU questionnaire. Proportions are also shown for two key demographic items—household language and English ability—that were included on the NATES questionnaire but not the NRFU. The table reports the percentage change in each proportion—defined as the change attributable to nonresponse adjustments, divided by the selection-weighted proportion—and a *t* test of the significance of the change. Although the NRFU follow-up sample is not used in this analysis, it is conducted at the household level (i.e. using only the data for the “first, most convenient” responder for households that returned more than one mailed questionnaire) for comparability with the analysis in the previous section.

**Table 2.3. Percentage distribution of NATES respondents, by weighting type and selected reported characteristics: 2013**

Selected reported characteristic	Selection-weighted percent	Nonresponse-adjusted percent	Change in estimate	Percent change in estimate	<i>t</i> test of change in estimate
Completed an apprenticeship					
Yes	8.0	8.1	0.1	1.3	0.7
No	92.0	91.9	-0.1	-0.1	-0.7
Completed other work-related training					
Yes	35.4	35.1	-0.3	-0.8	-1.4
No	64.6	64.9	0.3	0.4	1.4
Speaks a language other than English at home					
Yes	16.3	17.4	1.1	6.8	6.7*
No	83.7	82.6	-1.1	-1.3	-6.7*
English ability					
Speaks well or very well	89.1	89.2	0.0	0.1	0.1
Speaks not very well or not at all	10.9	10.8	0.0	-0.4	-0.1

\**p* < .001.

NOTE: Percentages represent the proportion of NATES respondents with the specified characteristic. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding. Standard errors of the selection-weighted and nonresponse-adjusted proportions, as well as for the change in the estimate, are shown in table D.3 of appendix D.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

It is clear that the nonresponse adjustments lead to very little change in the proportions for these items. The largest change is in the household language item, for which the nonresponse adjustments increase the estimated proportion of the population that speaks a language other than English at home by about 1.1 percentage points. This is also the only item for which the difference is statistically significant.

As noted above, because the amount of nonresponse bias in these items is unknown, the effectiveness of the adjustments at correcting for possible bias in these estimates cannot be known with certainty. However, to the extent that bias exists in these estimates, the fact that the adjustments lead to little change in the estimates suggests a minimal impact on the amount of nonresponse bias.<sup>14</sup> As a rough approximation of the extent to which this approach may underestimate nonresponse bias, it can be noted that for the three key items of interest in table 2.2 (certifications/licenses, certificates, and college enrollment), changes in the estimate due to nonresponse adjustment accounted for approximately 20 to 40 percent of the observed bias.

## 2.4 Summary and Discussion of Results

This chapter addressed the report's first three research questions by evaluating unit nonresponse bias in key NATES proportions both before and after nonresponse adjustment. The key finding from the analysis reported in section 2.1 was that bias existed in auxiliary variables when proportions for respondent households were estimated using selection weights, but was reduced when those same proportions were estimated using nonresponse-adjusted weights. In other words, the nonresponse adjustments reduced bias in the composition of the respondent pool, as measured by auxiliary variables. As noted above, because many of these auxiliary variables were identified by the household CHAID model as being predictive of household nonresponse, and were therefore used to define the household nonresponse adjustment cells, this finding is consistent with expectations. A secondary finding was that the selection-weighted proportions in the "missing" category of auxiliary variables were consistently lower in the respondent group than in the overall eligible sample, which implies that missing auxiliary data are predictive of household nonresponse.

Section 2.2 found that, when proportions were generated using nonresponse-adjusted weights, there was no statistically significant bias in the estimated prevalence of three key survey items measured by both questionnaires: certifications/licenses, certificates, and college enrollment. However, bias in many of the key demographic proportions remained statistically significant and substantively important, although this remaining bias could be removed through the use of a poststratification or raking adjustment. Thus, there appears to be little or no nonignorable bias in the proportions that could be evaluated using the NRFU.

While bias could not be estimated directly for the apprenticeship and training items, section 2.3 found that the nonresponse adjustments led to very small changes in the associated proportions. This is suggestive of limited effectiveness in reducing nonresponse bias; however, given the fact that no statistically significant bias was found in the selection-weighted certification/license and certificate proportions, it is possible that there was also little or no bias in the selection-weighted apprenticeship and training proportions.

---

<sup>14</sup> Similar results are obtained if this analysis is conducted using person-level observations and weights (not shown in tables). Specifically, when nonresponse-adjusted person-level estimates are compared to base-weighted person-level estimates, the estimates change minimally, and only the household language item shows a statistically significant change.

The methods used in sections 2.1 and 2.3—comparing the distribution of auxiliary variables between respondents and the sample, and comparing selection-weighted and nonresponse-adjusted key estimates, respectively—represent methods that will be able to be used for nonresponse bias analysis in future, full-scale ATES administrations. In contrast, the NRFU was a one-time feature of the NATES pilot study and is not planned to be repeated in the near future. By comparing the results discussed in section 2.2 to those discussed in sections 2.1 and 2.3, it is possible to gain insight into the extent to which some of the methods that are available in the absence of an NRFU can accurately indicate the presence or absence of bias in key survey estimates.

As discussed in section 2.1, the distributions of nearly all available auxiliary variables showed statistically significant and substantively important bias prior to nonresponse adjustment, but much of this bias was removed by the adjustment. Taken in isolation, these results would have suggested that there was a high risk of nonresponse bias in the NATES estimates, but that this risk was substantially mitigated once the weights were adjusted for nonresponse. However, the analysis using the NRFU found that, while substantial bias existed in demographic proportions, the estimated prevalence of certifications/licenses and certificates did not show statistically significant bias, while the estimated college enrollment rate showed statistically significant bias of approximately 1.5 percentage points. The nonresponse adjustments led to limited changes in these key proportions of interest, although the bias in the college enrollment rate was no longer statistically significant after adjustment. Thus, the NRFU results suggest that bias in the key proportions is likely to be limited prior to adjustment; but that, to the extent that it does exist, it is unlikely to be substantially mitigated through the use of weighting adjustments. Consequently, a nonresponse bias analysis using only the auxiliary variables available in or linked to the frame would be likely to overestimate the risk of bias prior to adjustment, but also to overestimate the extent to which the adjustments remove bias in key survey estimates. In contrast, comparing the results in sections 2.2 and 2.3, it is evident that the change in estimates attributable to nonresponse adjustment may substantially *underestimate* the magnitude of bias prior to adjustment, insofar as the NRFU results suggest that the nonresponse adjustment accounts for only a fraction of the observed bias.

The analysis that follows in chapter 3 is intended to evaluate two possible strategies for using auxiliary data to minimize nonresponse bias in future ATES administrations. The first would be to improve the effectiveness of the nonresponse adjustment process by augmenting the availability of data for demographic variables used in the generation of nonresponse-adjusted weights. The second would be to increase response rates among households with characteristics that are associated with nonresponse. As explained in detail in chapter 3, the effectiveness of either of these strategies at mitigating nonresponse bias would depend on the extent to which characteristics known for both respondents and nonrespondents are associated with responses to the items of interest. For this reason, chapter 3 focuses on evaluating the strength of the associations between the variables available in the NATES sampling frame and responses to the key survey items.

## Chapter 3: Evaluation of Strategies for Reducing Unit Nonresponse Bias

This chapter considers two potential approaches to addressing nonresponse bias in household studies of adults. The first would be to improve the effectiveness of the nonresponse adjustment process, which, as shown in chapter 2, only moderately reduced the estimated nonresponse bias in the NATES estimates. As discussed in chapter 1, that process allocates sample members to adjustment cells defined by characteristics found to be predictive of nonresponse. These characteristics must be available in or linked to the sampling frame, and high missing rates for these auxiliary variables can undermine the effectiveness of the adjustment process at correcting for nonresponse bias. In the case of NATES, many of the auxiliary variables used in the nonresponse adjustment process were missing for large proportions of the sample (in some cases, over 40 percent). This raises the question of whether the adjustments could be improved by increasing the coverage of these variables; specifically, by appending supplemental data purchased from other commercial sources.<sup>15</sup> From a statistical perspective, the answer to this question depends on whether, among households for which these variables *are* available, they are significantly and consistently associated with responses to the NATES items. If this is found to be the case, and it is assumed that the same relationships exist among households for which these auxiliary variables are missing, then in theory it may be possible to improve the nonresponse adjustments by obtaining data on these variables for a larger proportion of the sample. However, if these auxiliary variables are not consistently related to key survey variables, then there is little reason to expect that the effectiveness of the nonresponse adjustment process would substantially improve if the coverage of the auxiliary variables was augmented.

A second possible strategy for reducing nonresponse bias would be to increase response rates, particularly among harder-to-reach population subgroups. NCES household surveys already incorporate a number of techniques to maximize overall response rates, including the use of incentives and follow-up mailings. The question of interest is whether, from the perspective of reducing nonresponse bias, it would be worthwhile to target particularly intensive versions of these techniques—for example, higher-than-normal incentives or phone follow-ups to supplement the standard mail follow-ups—at certain demographic subgroups. Again, the effectiveness of this approach would depend on the strength of the associations between demographic variables linked to the sampling frame and responses to the NATES items. An increase in response rates would be expected to reduce nonresponse bias only if the additional respondents were likely to provide a different distribution of responses than existing respondents. Consequently, in choosing variables that could potentially be used to define targeting criteria, it is

---

<sup>15</sup> West, Wagner, Hubbard, and Gu (2013) provide an example of linking demographic data from multiple commercial vendors for a single study. For some vendors, missing data rates on appended auxiliary variables can also be affected by the order in which the sample is matched to the available databases, and on whether cases that are matched to one source are then also matched to subsequent sources.

necessary to identify variables that are significantly associated both with the propensity to respond and with the actual response provided to key items.

To evaluate each of these potential approaches, section 3.1 analyzes the associations between auxiliary variables available for NATES and responses to the key survey items.

Section 3.2 evaluates a variant of the second approach, in which the sampling frame would be used to target households based not on their specific characteristics, but rather on *how much* auxiliary data are available for the household. The impetus for this analysis was the finding that, for many of the auxiliary variables evaluated in chapter 2 (see table 2.1), missing rates in the were significantly lower among respondents than in the overall eligible sample. This implies that households that are missing auxiliary data for particular characteristics were less likely to respond. The fact that missing auxiliary data appears to be associated with nonresponse raises the question of whether households with missing auxiliary data are different from other households in a way that would affect their propensity to report a credential. If so, then increased response rates among households with missing auxiliary data would be expected to reduce nonresponse bias in estimates of credential prevalence—which implies that these households may be good candidates for the targeting of incentives or follow-up. For this reason, section 3.2 first evaluates the association between the number of missing auxiliary variables and response rates to the mailed NATES questionnaire and then between the number of missing auxiliary variables and responses to the key survey items.

Section 3.3 follows with a discussion of the implications of the results presented in sections 3.1 and 3.2, as they pertain to the two key research questions addressed by this chapter.

Since data on auxiliary variables are available only at the household level, and data are combined from NATES and NRFU respondents, both analyses use household-level data and weights.<sup>16</sup> Standard errors are calculated using the jackknife method with 70 replicate weights.

### 3.1 Auxiliary Variables and Responses to Key Items

In this section, binomial logistic regression models<sup>17</sup> are estimated for the response to five key items from the NATES questionnaire:

- Has certification or license (Q4)
- Has educational certificate (Q20)
- Completed an apprenticeship (Q29)<sup>18</sup>
- Currently enrolled in college classes (Q35)
- Completed work-related training (Q46)

---

<sup>16</sup>As in chapter 2, the NATES person-level survey responses were converted to a household-level file by keeping only the responses from the first NATES respondent to return a questionnaire in each household. Household-level selection weights were then applied to the remaining NATES observations. See section E.5 of appendix E for a detailed description of this approach and its rationale.

<sup>17</sup> Although not reported in this chapter, bivariate analysis of the relationship between auxiliary variables and credential reporting shows similar results to the multivariate analysis reported here—namely, that few significant relationships exist.

<sup>18</sup> The responses to the apprenticeship item were recoded to a binary dependent variable.

For the certification/license, educational certificate, and college enrollment items, the regressions are estimated using selection-weighted data from respondents to the mailed NATES questionnaire *and* respondents to the in-person NRFU questionnaire. Because, as discussed in chapter 1, the combined selection-weighted NATES and NRFU sample is roughly representative of the entire eligible population, the inclusion of the NRFU respondents allows a more accurate determination of which auxiliary variables are associated with the prevalence of these credentials in the population as a whole. The apprenticeship and work-related training items were not included in the NRFU questionnaire; consequently, these regressions are estimated using only selection-weighted data from respondents to the mailed NATES questionnaire.

The initial set of candidate independent variables for each regression in this section consisted of the auxiliary variables that were found to be predictive of response status by the household-level CHAID model: address route type; the availability of a phone number in the sampling frame; home tenure (own or rent); household income; the number of adults in the household; and the age, educational attainment, ethnicity, and gender of the head of the household. These were the variables used to define the adjustment cells in the household-level nonresponse adjustment process. In addition, for the three regressions estimated using both NATES and NRFU data, an indicator variable for NRFU respondents is included to control for nonresponse and/or mode effects.<sup>19</sup> Because of the large number of candidate independent variables, a modified version of the model-building process proposed by Heeringa, West, and Berglund (2010) was used independently for each of the five models:

1. Variance inflation factors (VIFs) were calculated for each auxiliary variable and the NRFU indicator to evaluate the potential effect of multicollinearity. No variables were dropped due to multicollinearity, but the reference categories for the age and income variables were changed to reduce the VIFs.
2. For each auxiliary variable, the strength of its bivariate association with the dependent variable was evaluated using a design-adjusted Pearson test of independence. Variables with a  $p$  value of .25 or below were selected for inclusion in the multivariate model.
3. For the three regressions that included NRFU observations, potential interactions between the indicator for NRFU respondents and each of the auxiliary variables selected in step 2 were assessed. A set of preliminary logistic regression models was estimated, each of which included as independent variables the auxiliary variables chosen in step 2, the NRFU indicator, and an interaction between the NRFU indicator and one of the auxiliary variables. Any interaction found to be significant at the .05 level using a joint design-adjusted Wald test was included in the final model.
4. For the regressions that did not include NRFU observations, the final multivariate model consisted of the variables selected in step 2. For those that did include NRFU observations, the

---

<sup>19</sup> Because the NRFU respondents were, by definition, NATES nonrespondents, and the auxiliary variables used in the models are known to be related to NATES response status (section 2.1 of chapter 2), the inclusion of the NRFU helps to prevent bias in the regression parameters that could be caused by relationships between the mode of response and the reporting of key credentials.

final multivariate model consisted of the variables selected in step 2 and the indicator for the NRFU observations (as no interactions were found to be significant in step 3).<sup>20</sup>

Because the question of interest in this analysis is whether auxiliary variables are predictive of credential attainment and participation in education and training in households for which those characteristics are available, cases with missing data for any independent variable are dropped from the regressions. Refer to table E.4 in appendix E for missing data rates for the auxiliary variables, including those used as independent variables in this analysis. Cases with missing data for the dependent variable were also excluded; unweighted missing data rates were 3.6 percent for certifications/licenses, 3.0 percent for certificates, 2.9 percent for college enrollment, 3.3 percent for apprenticeships, and 2.5 percent for trainings (not shown in tables).

All of the independent variables in these regressions are categorical in nature. Each is therefore recoded to a set of binary indicator variables—one for each category, minus a base category that serves as a reference group—for the purpose of the regressions. Accordingly, with the exception of the base category, each category of each independent variable has its own coefficient in the regression tables. The coefficients are reported as “odds ratios,” which can be interpreted as the odds of reporting the credential among respondents in the specified category divided by the odds of reporting the credential among respondents in the base category.<sup>21</sup> For example, in the logistic regression for the response to the certification/license item (see table 3.2 below), the “high rise” category is the base category for the route type variable and the “rural” category has an odds ratio of 1.8. The interpretation of this result is that the odds of a household with a rural address reporting a certification/license are 1.8 times as high as the odds of a household with a high-rise address reporting a certification/license. Odds ratios below 1 for a particular category imply lower odds of reporting the credential than the reference group, while odds ratios above 1 imply higher odds of reporting the credential. The statistical significance of each individual odds ratio is assessed using a *t* test, which indicates whether the odds ratio is significantly greater than or less than 1.

In this particular analysis, the key outcome of interest is the strength of the *overall* association between each categorical auxiliary variable and the propensity to report a credential or type of training. Therefore, for each independent variable, the tables also report the results of a design-adjusted Wald *F* test for the joint significance of the coefficients on all of that variable’s categories. A significant Wald *F* statistic value suggests that the auxiliary variable in question is significantly associated with the response to the item. Therefore, in this section, references to the statistical significance of a particular auxiliary variable

---

<sup>20</sup> For the certification/license regression, the number of adults was selected as an independent variable based on this process, but was dropped from the final regression because one of its categories perfectly predicted the dependent variable. Specifically, none of the respondents with 8 adults in the household reported a certification/license. The route type was dropped from the final college enrollment regression for the same reason—none of the households with a rural address type reported being enrolled in college courses. When perfect prediction occurs, odds ratios cannot be calculated because unique maximum likelihood estimates do not exist (SAS Institute 2011). As a robustness check, the certification/license regression (table 3.2) was run using a version of the number of adults variable with the “7 adults” and “8 adults” categories collapsed, which led to no substantial change in the results. Similarly, the college enrollment regression (table 3.4) was rerun using a version of the route type variable with the “rural route” and “P.O. Box” categories collapsed; the age of the head of the household remained statistically significant while the home tenure variable was no longer statistically significant.

<sup>21</sup> The “odds” of reporting a credential refers to the probability of responding “yes” to the item divided by the probability of responding “no.”



(e.g., the educational attainment variable) refer to the results of the joint Wald significance test for that variable.

Table 3.1 summarizes the regression results by identifying the auxiliary variables that were found to be significantly associated with the propensity to report each estimate of interest. Tables 3.2 through 3.6 report detailed regression results from the final models for each credential or training item. In the regression for the certification/license item (table 3.2), only the address route type is statistically significant. No auxiliary variables are significant predictors of certificate reporting (table 3.3). Home tenure and the age of the head of household are significant predictors of college enrollment (table 3.4). Neither education nor ethnicity is a significant predictor of the propensity to report an apprenticeship (table 3.5). Finally, household income, the age of the head of the household, and the educational attainment of the head of the household are associated with the propensity to report other work-related training (table 3.6).

**Table 3.1. Associations between auxiliary variables and reporting key estimates, by key survey item and characteristic: 2013**

Characteristic	Associated with reporting of				
	Certification or license	Certificate	College enrollment	Apprenticeship	Training
Route type	X	†	†	†	†
Phone number available in sampling frame		†		†	†
Home tenure		†	X	†	
Age of the head of household	†		X	†	X
Household income		†	†	†	X
Number of adults in household	†	†	†	†	
Education of the head of household			†		X
Race/ethnicity of the head of household	†		†		
Gender of the head of household		†		†	†

† Not applicable (characteristic was not included in the final regression model for the specified credential).

NOTE: An "X" indicates that a statistically significant association existed between the specified characteristic and the propensity to report the credential or training, as determined by an adjusted joint Wald test on logistic regression results. The route type variable is taken from the U.S. Postal Service's Computerized Delivery Sequence (CDS) file. Data for other independent variables was purchased from commercial vendors and appended to the NATES sampling frame. This commercial data is not equivalent to respondents' self-reported data, and its accuracy may vary. The "head of the household" may not be the same individual as the NATES or NRFU respondent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

In summary, the age of the head of household, based on data provided by the sampling frame vendor, is significantly associated with responses to two of the key survey items (college enrollment and training). No other auxiliary variables are statistically significant in more than one of the regressions.

These results provide an indication of why the nonresponse adjustments led, in general, to only moderate changes in the key survey estimates. Specifically, the limited effectiveness of the nonresponse adjustments appears to be driven not only by high missing rates for the auxiliary variables, but by a limited association between those variables and credential reporting among households for which the variables are not missing.

**Table 3.2. Odds ratios of reporting a certification or license, by respondent sample and selected characteristics: 2013**

Respondent sample or selected characteristic	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic	Wald F statistic	Wald p value
Respondent sample							
NATES respondents	†	†	†	†	†	0.0	0.838
NRFU respondents	1.0	0.15	0.8	1.4	0.2		
Route type							
High rise	†	†	†	†	†	4.8*	0.005
Street	1.0	0.20	0.6	1.4	-0.2		
Rural	1.8	0.54	1.0	3.3	2.0*		
P.O. box	0.6	0.31	0.2	1.7	-0.9		
Phone number available in sampling frame							
Yes	†	†	†	†	†	1.4	0.245
No	1.1	0.13	0.9	1.4	1.2		
Home tenure							
Own	†	†	†	†	†	0.8	0.370
Rent	0.9	0.14	0.6	1.2	-0.9		
Household income							
\$0–\$10,000	0.9	0.53	0.2	3.0	-0.3	1.1	0.348
\$10,001–\$20,000	0.4	0.17	0.2	0.9	-2.2*		
\$20,001–\$30,000	0.5	0.16	0.3	0.9	-2.2*		
\$30,001–\$40,000	0.6	0.19	0.3	1.1	-1.6		
\$40,001–\$50,000	0.7	0.20	0.4	1.3	-1.2		
\$50,001–\$60,000	0.8	0.24	0.4	1.4	-0.8		
\$60,001–\$75,000	0.8	0.21	0.5	1.4	-0.8		
\$75,001–\$100,000	0.9	0.22	0.5	1.4	-0.5		
\$100,000–\$150,000	0.9	0.21	0.5	1.4	-0.6		
\$150,001+	†	†	†	†	†		
Education of the head of household							
Less than high school diploma	†	†	†	†	†	1.6	0.198
High school diploma	0.9	0.17	0.6	1.3	-0.4		
Some college	1.1	0.20	0.7	1.6	0.4		
Bachelor's degree	1.3	0.26	0.9	2.0	1.5		
Graduate degree	1.4	0.33	0.9	2.3	1.6		
Gender of the head of household							
Female	†	†	†	†	†	0.0	0.944
Male	1.0	0.15	0.8	1.4	0.1		

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

NOTE: Unweighted sample size for this model is approximately 3,140. Approximately 2,390 observations are excluded due to missing data for an independent or dependent variable. "Respondent sample" indicates whether the observation came from a respondent to the mailed NATES questionnaire or the in-person NRFU interview. The route type variable is taken from the U.S. Postal Service's Computerized Delivery Sequence (CDS) file. Data for other independent variable was purchased from commercial vendors and appended to the NATES sampling frame. This commercial data is not equivalent to respondents' self-reported data, and its accuracy may vary. The "head of the household" may not be the same individual as the NATES or NRFU respondent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table 3.3. Odds ratios of reporting a certificate, by respondent sample and selected characteristics: 2013**

Respondent sample or selected characteristic	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic	Wald F statistic	Wald p value
Respondent sample							
NATES respondents	†	†	†	†	†	3.5	0.068
NRFU respondents	1.5	0.30	1.0	2.2	1.9		
Age of the head of household							
18–24	0.4	0.33	0.1	2.0	-1.1	2.4	0.050 <sup>1</sup>
25–34	0.5	0.18	0.2	1.0	-1.9		
35–44	0.6	0.18	0.3	1.1	-1.7		
45–54	0.8	0.21	0.5	1.4	-0.7		
55–65	0.5	0.13	0.3	0.8	-2.8*		
Over 65	†	†	†	†	†		
Education of the head of household							
Less than high school diploma	†	†	†	†	†	2.3	0.068
High school diploma	1.0	0.31	0.5	1.8	-0.1		
Some college	1.0	0.31	0.5	1.8	-0.2		
Bachelor's degree	0.6	0.21	0.3	1.2	-1.5		
Graduate degree	0.7	0.24	0.4	1.4	-1.0		
Race/ethnicity of the head of household <sup>2</sup>							
White	†	†	†	†	†	0.3	0.876
Black	0.8	0.24	0.5	1.5	-0.6		
Hispanic	1.0	0.31	0.6	1.9	0.1		
Asian/Pacific Islander	0.7	0.26	0.4	1.5	-0.9		
Other	1.0	0.24	0.6	1.6	0.0		

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

<sup>1</sup> The unrounded  $p$  value is above .05; therefore, the result is not treated as statistically significant.

<sup>2</sup> Race/ethnicity categories were based on the vendor auxiliary variable "Ethnicity" which combined race and ethnicity into one variable. "White" included these categories from the vendor's frame: Czech, Dutch, Eastern European, English, French, German, Greek, Irish, Italian, Jewish, Middle Eastern, Polish, Portuguese, Russian, Scandinavian, Scottish, Swiss, Ukrainian, and Western European. "Black" included African and African American. "Hispanic" included Hispanic. "Asian or Pacific Islander" included Asian, Chinese, Hawaiian, Indonesian, Japanese, Korean, Polynesian, and Vietnamese. "Other" included Miscellaneous Other, Native American, and unknown. Note that the auxiliary variable does not provide information on possible interactions between race and Hispanic ethnicity; therefore, some non-Hispanic racial groups may include Hispanics.

NOTE: Unweighted sample size for this model is approximately 2,690. Approximately 2,840 observations are excluded due to missing data for an independent or dependent variable. "Respondent sample" indicates whether the observation came from a respondent to the mailed NATES questionnaire or the in-person NRFU interview. Data for all other independent variables was purchased from commercial vendors and appended to the NATES sampling frame. This commercial data is not equivalent to respondents' self-reported data, and its accuracy may vary. The "head of the household" may not be the same individual as the NATES or NRFU respondent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey, 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table 3.4. Odds ratios of reporting college enrollment, by respondent sample and selected characteristics: 2013**

Respondent sample or selected characteristic	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic	Wald F statistic	Wald p value
Respondent sample							
NATES respondents	†	†	†	†	†	0.9	0.337
NRFU respondents	1.3	0.31	0.8	2.1	1.0		
Phone number available in sampling frame							
Yes	†	†	†	†	†	0.0	0.864
No	1.0	0.24	0.7	1.6	0.2		
Home tenure							
Own	†	†	†	†	†	5.5*	0.021
Rent	1.8	0.42	1.1	2.8	2.4*		
Age of the head of household							
18–24	1.3	0.84	0.4	4.7	0.4	4.3*	0.002
25–34	1.0	0.41	0.4	2.3	0.0		
35–44	0.8	0.31	0.4	1.7	-0.6		
45–54	0.7	0.21	0.4	1.3	-1.1		
55–65	0.3	0.11	0.2	0.6	-3.4*		
Over 65	†	†	†	†	†		
Gender of the head of household							
Female	†	†	†	†	†	1.2	0.277
Male	1.3	0.26	0.8	1.9	1.1		

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

NOTE: Unweighted sample size for this model is approximately 3,220. Approximately 2,310 observations are excluded due to missing data on an independent or dependent variable. "Respondent sample" indicates whether the observation came from a respondent to the mailed NATES questionnaire or the in-person NRFU interview. Data for all other independent variables was purchased from commercial vendors and appended to the NATES sampling frame. This commercial data is not equivalent to respondents' self-reported data, and its accuracy may vary. The "head of the household" may not be the same individual as the NATES or NRFU respondent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table 3.5. Odds ratios of reporting an apprenticeship, by selected characteristics: 2013**

Selected characteristic	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic	Wald F statistic	Wald p value
Education of the head of household							
Less than high school diploma	†	†	†	†	†	1.8	0.133
High school diploma	1.3	0.49	0.6	2.8	0.8		
Some college	1.4	0.46	0.7	2.7	0.9		
Bachelor's degree	0.7	0.25	0.3	1.4	-1.1		
Graduate degree	1.1	0.37	0.5	2.1	0.2		
Race/ethnicity of the head of household <sup>1</sup>							
White	†	†	†	†	†	1.1	0.365
Black	1.4	0.48	0.7	2.7	0.9		
Hispanic	1.9	0.82	0.8	4.5	1.6		
Asian/Pacific Islander	1.6	0.48	0.9	2.9	1.5		
Other	0.8	0.36	0.4	2.0	-0.4		

† Not applicable (reference group for odds ratio).

<sup>1</sup> Race/ethnicity categories were based on the vendor auxiliary variable "Ethnicity" which combined race and ethnicity into one variable. "White" included these categories from the vendor's frame: Czech, Dutch, Eastern European, English, French, German, Greek, Irish, Italian, Jewish, Middle Eastern, Polish, Portuguese, Russian, Scandinavian, Scottish, Swiss, Ukrainian, and Western European. "Black" included African and African American. "Hispanic" included Hispanic. "Asian or Pacific Islander" included Asian, Chinese, Hawaiian, Indonesian, Japanese, Korean, Polynesian, and Vietnamese. "Other" included Miscellaneous Other, Native American, and unknown. Note that the auxiliary variable does not provide information on possible interactions between race and Hispanic ethnicity; therefore, some non-Hispanic racial groups may include Hispanics.

NOTE: Unweighted sample size for this model is approximately 2,720. Approximately 1,630 observations are excluded due to missing data for an independent or dependent variable. Data for all independent variables was purchased from commercial vendors and appended to the NATES sampling frame. This commercial data is not equivalent to respondents' self-reported data, and its accuracy may vary. The "head of the household" may not be the same individual as the NATES or NRFU respondent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table 3.6. Odds ratios of reporting other work-related training, by selected characteristics: 2013**

Selected characteristic	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic	Wald F statistic	Wald p value
Home tenure							
Own	†	†	†	†	†	1.7	0.197
Rent	0.8	0.15	0.5	1.1	-1.3		
Age of the head of household							
18–24	1.2	0.49	0.5	2.7	0.5	2.6*	0.032
25–34	1.9	0.59	1.0	3.5	2.1*		
35–44	1.6	0.40	0.9	2.6	1.7		
45–54	1.5	0.38	0.9	2.5	1.6		
55–65	1.0	0.27	0.6	1.7	0.1		
Over 65	†	†	†	†	†		
Household income							
\$0–\$10,000	0.7	0.67	0.1	4.7	-0.4	2.3*	0.029
\$10,001–\$20,000	0.4	0.13	0.2	0.7	-2.8*		
\$20,001–\$30,000	0.3	0.10	0.2	0.6	-3.8*		
\$30,001–\$40,000	0.5	0.13	0.3	0.8	-2.7*		
\$40,001–\$50,000	0.5	0.12	0.3	0.8	-3.0*		
\$50,001–\$60,000	0.7	0.19	0.4	1.3	-1.1		
\$60,001–\$75,000	0.6	0.15	0.4	1.0	-2.1*		
\$75,001–\$100,000	0.7	0.16	0.5	1.1	-1.5		
\$100,000–\$150,000	0.7	0.12	0.5	0.9	-2.3*		
\$150,001+	†	†	†	†	†		
Number of adults in household							
1	†	†	†	†	†	0.8	0.609
2	1.0	0.17	0.7	1.4	-0.2		
3	0.9	0.15	0.7	1.3	-0.5		
4	1.0	0.23	0.6	1.6	-0.1		
5	0.6	0.17	0.3	1.0	-1.9		
6	0.6	0.26	0.3	1.4	-1.2		
7	1.6	1.05	0.5	5.9	0.8		
Education of the head of household							
Less than high school diploma	†	†	†	†	†	4.5*	0.003
High school diploma	0.8	0.32	0.3	1.7	-0.7		
Some college	1.3	0.50	0.6	2.8	0.6		
Bachelor's degree	1.5	0.62	0.7	3.4	1.0		
Graduate degree	1.2	0.51	0.5	2.8	0.5		

See notes at end of table.

**Table 3.6. Odds ratios of reporting other work-related training, by selected characteristics: 2013—Continued**

Selected characteristic	Odds ratio	Standard error	95% confidence interval: upper bound	95% confidence interval: lower bound	t statistic	Wald F statistic	p value of Wald F
Race/ethnicity of the head of household <sup>†</sup>							
White	†	†	†	†	†	1.4	0.260
Black	1.0	0.19	0.7	1.5	0.3		
Hispanic	0.7	0.15	0.4	1.1	-1.7		
Asian/Pacific Islander	0.6	0.14	0.4	1.0	-2.1*		
Other	0.9	0.20	0.6	1.4	-0.3		

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

<sup>†</sup> Race/ethnicity categories were based on the vendor auxiliary variable “Ethnicity” which combined race and ethnicity into one variable. “White” included these categories from the vendor’s frame: Czech, Dutch, Eastern European, English, French, German, Greek, Irish, Italian, Jewish, Middle Eastern, Polish, Portuguese, Russian, Scandinavian, Scottish, Swiss, Ukrainian, and Western European. “Black” included African and African American. “Hispanic” included Hispanic. “Asian or Pacific Islander” included Asian, Chinese, Hawaiian, Indonesian, Japanese, Korean, Polynesian, and Vietnamese. “Other” included Miscellaneous Other, Native American, and unknown. Note that the auxiliary variable does not provide information on possible interactions between race and Hispanic ethnicity; therefore, some non-Hispanic racial groups may include Hispanics.

NOTE: Unweighted sample size for this model is approximately 2,210. Approximately 2,140 observations are excluded due to missing data on an independent or dependent variable. Data for all independent variables was purchased from commercial vendors and appended to the NATES sampling frame. This commercial data is not equivalent to respondents’ self-reported data, and its accuracy may vary. The “head of the household” may not be the same individual as the NATES or NRFU respondent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

### 3.2 Availability of Auxiliary Variables and Responses to Key Items

For the purpose of this analysis, key auxiliary variables are defined as those for which a statistically significant difference in the missing proportion was found between respondents and the total population, based on the analysis reported in table 2.1. These variables include phone number; home tenure; dwelling type; household income; the number of adults in the household; and the ethnicity, educational attainment, age, gender, and marital status of the head of the household. These variables were chosen because this result implies that missing data for any one of these auxiliary variables is predictive of nonresponse.

Table 3.7 reports selection-weighted household-level unit response rates to the mailed NATES questionnaire by the number of key auxiliary variables for which data were missing for the household. Counts of missing variables were used because the research question of interest was whether missing data in general (as opposed to missing data in a particular auxiliary variable) is related to key survey variables. The response rate among households missing one key auxiliary variable is over 8 percentage points lower than that among households that are missing no key auxiliary variables, although still slightly above the overall household-level NATES response rate of 65 percent. The response rate falls below 65 percent when two auxiliary variables are missing and continues to decrease as the number of missing auxiliary variables increases. These results further illustrate that missing auxiliary data are predictive of unit

nonresponse—which is to be expected, as households for which the commercial vendor was unable to obtain complete data are likely to be households that, in general, are difficult to reach.<sup>22</sup>

**Table 3.7. Percentage distribution of NATES eligible sample and weighted response rate to mailed NATES questionnaire, by number of key auxiliary variables missing for household: 2013**

Number of key auxiliary variables missing for household	Percent of NATES eligible sample	Unit response rate to mailed NATES questionnaire
0	22.1	75.0
1	23.8	66.7
2	15.1	64.1
3	12.7	62.5
4	7.6	59.1
5 or more	18.8	58.2

NOTE: Percentages represent the weighted proportion of eligible NATES households missing the specified number of key auxiliary variables. Response rates represent the weighted proportion of eligible NATES households that returned at least one questionnaire. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

With this established, the analysis reported in the remainder of this section proceeds in a similar manner as the analysis in section 3.1. Logistic regressions are estimated with responses to the same five key NATES items as the dependent variables. The independent variable in each regression is a binary indicator coded as “1 or fewer auxiliary variables missing” and “2 or more auxiliary variables missing.” Two was chosen as the cutoff because it is the threshold at which response rates fall below the overall household response rate and because slightly more than half of eligible households have two or more missing auxiliary variables. Consequently, it is assumed that this would make a reasonable and feasible criterion for targeting incentives, follow-up efforts, or other nonresponse conversion methods.<sup>23</sup>

As in section 3.1, the regressions for the certification/license, certificate, and college enrollment item are estimated using data from NATES and NRFU respondents, because the combined NATES and NRFU respondents are more representative of the target population than the NATES respondents alone. These

<sup>22</sup> It is also worth noting that address ineligibility rates tend to increase as the number of missing auxiliary variables increase. In NATES, the weighted address ineligibility rate for households with no missing auxiliary variables was 2.9 percent, compared to 26.0 percent for households with 5 or more (not shown in tables).

<sup>23</sup> Because the largest drop-off in response rates occurred when the missing count increased from zero to one, an alternative approach would be to code the independent variable as “0 auxiliary variables missing” or “1 or more auxiliary variables missing.” This approach was tested, but rejected for two reasons. First, the “1 or more” category would encompass over 78 percent of eligible households (unweighted, not shown in tables), meaning it would be operationally infeasible to intensively target all households in this category. Second, when the regressions were tested with the variables coded in this way (not shown in tables), the observed relationships between the missing count and credential reporting were weaker than the results reported here. This suggests that targeting the approximately 55 percent of households with two or more missing auxiliary variables—the approach reflected by using two as the cutoff for the binary coding—would be a more efficient approach to the reduction of nonresponse bias. A second alternative approach would be to treat the missing count as a continuous independent variable. When this approach was tested (not shown in tables), the sign and significance of the relationships between the missing count and credential reporting were the same as in the results presented in this chapter.



regressions include an indicator variable for NRFU observations as a control for mode effects and/or nonresponse. Interactions between the NRFU indicator and the missing count were tested; the interaction was significant at the .05 level in the regression for certificates and was therefore retained in that model. The regressions for the apprenticeship and work-related training items are estimated using only NATES data and thus exclude the NRFU indicator. Because the missing count variable is categorical, the same interpretation of the odds ratios and significance tests discussed in section 3.1 applies.

Tables 3.8 through 3.12 provide logistic regression results for each item. The regression for certifications/licenses (table 3.8) indicates that respondents with two or more missing auxiliary variables had significantly lower odds of reporting a certification or license. If households with a large number of missing auxiliary variables are less likely to respond to the mailed questionnaire, and are less likely to hold a certification/license, then estimates based on responses to the mailed questionnaire would be expected to overestimate the prevalence of certifications/licenses relative to the true population prevalence. Consistent with this expectation, chapter 2 found a positive (though not statistically significant) bias in the proportion of NATES respondents with a certification/license.

In the regression for certificates (table 3.9), the odds ratio for the missing count is insignificant at the .05 level, but the odds ratio for the interaction between the NRFU indicator and the missing count is significant. This suggests that the relationship between the missing count and certificate reporting is different for NATES and NRFU respondents. Upon further examination, it appears that among NRFU respondents only, individuals with two or more missing auxiliary variables had significantly lower odds of reporting a certificate than individuals with less than two missing auxiliary variables. This association was not observed among NATES respondents (not shown in tables). Thus, among nonrespondents to NATES, the relationship between missing auxiliary data and credential reporting is similar for certificates as for certifications and licenses; that is, households missing data for two or more auxiliary variables were less likely to report a certificate.

The regression for college enrollment (table 3.10) finds that respondents with two or more missing auxiliary variables had significantly higher odds of reporting enrollment in college classes. This implies that estimates based on responses to the mailed questionnaire would be expected to underestimate the prevalence of college enrollment—which, again, is consistent with the direction of the bias reported in chapter 2.

Finally, based on the insignificant odds ratios, the propensity to report an apprenticeship (table 3.11) or work-related training (table 3.12) appears to be largely independent of the number of missing auxiliary variables.

**Table 3.8. Odds ratios of reporting a certification or license, by respondent sample and number of missing auxiliary variables: 2013**

Respondent sample or number of missing auxiliary variables	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic
Respondent sample					
NATES respondents	†	†	†	†	†
NRFU respondents	0.9	0.10	0.7	1.1	-1.0
Number of missing auxiliary variables					
1 or fewer	†	†	†	†	†
2 or more	0.8	0.07	0.7	1.0	-2.2*

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

NOTE: Unweighted sample size for this model is approximately 5,330. Approximately 200 observations are excluded due to missing data on the dependent variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table 3.9. Odds ratios of reporting a certificate, by respondent sample and number of missing auxiliary variables: 2013**

Respondent sample or number of missing auxiliary variables	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic
Respondent sample					
NATES respondents	†	†	†	†	†
NRFU respondents	1.6	0.33	1.1	2.4	2.3*
Number of missing auxiliary variables					
1 or fewer	†	†	†	†	†
2 or more	1.1	0.15	0.9	1.5	0.8
Interaction term					
2 or more X NRFU respondents	0.6	0.14	0.3	0.9	-2.4*

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

NOTE: Unweighted sample size for this model is approximately 5,360. Approximately 170 observations are excluded due to missing data on the dependent variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table 3.10. Odds ratios of reporting college enrollment, by respondent sample and number of missing auxiliary variables: 2013**

Respondent sample or number of missing auxiliary variables	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic
Respondent sample					
NATES respondents	†	†	†	†	†
NRFU respondents	1.4	0.23	1.0	2.0	2.2*
Number of missing auxiliary variables					
1 or fewer	†	†	†	†	†
2 or more	1.4	0.24	1.0	2.0	2.2*

† Not applicable (reference group for odds ratio).

\* $p < .05$ .

NOTE: Unweighted sample size for this model is approximately 5,370. Approximately 160 observations are excluded due to missing data on the dependent variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table 3.11. Odds ratios of reporting an apprenticeship, by number of missing auxiliary variables: 2013**

Number of missing auxiliary variables	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic
1 or fewer	†	†	†	†	†
2 or more	1.1	0.203	0.7	1.6	0.4

† Not applicable (reference group for odds ratio).

NOTE: Unweighted sample size for this model is approximately 4,210. Approximately 140 observations are excluded due to missing data on the dependent variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table 3.12. Odds ratios of reporting work-related training, by number of missing auxiliary variables: 2013**

Number of missing auxiliary variables	Odds ratio	Standard error	95% confidence interval: lower bound	95% confidence interval: upper bound	t statistic
1 or fewer	†	†	†	†	†
2 or more	0.9	0.09	0.8	1.1	-0.6

† Not applicable (reference group for odds ratio).

NOTE: Unweighted sample size for this model is approximately 4,240. Approximately 110 observations are excluded due to missing data on the dependent variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

### 3.3 Discussion of Results

Section 3.1 evaluated the strength of the association between various demographic characteristics available in or linked to the sampling frame and the propensity to report a certification/license, a certificate, college enrollment, an apprenticeship, or work-related training. No auxiliary variable was significantly associated with responses to more than one key survey item, except the age of the head of the household, which was found to be significantly associated with the propensity to report a certificate, college enrollment, or training. This indicates that the limited efficacy of the nonresponse adjustments resulted not only from the auxiliary variables having high missing rates, but also from generally weak associations between these variables and responses to the items of interest.

The weakness of the observed associations between auxiliary variables and survey variables—even for auxiliary variables, like educational attainment, that would be expected to correlate more strongly with the credentials and training measured by NATES—may be attributable to several characteristics of commercial auxiliary data. The first is that the data are available at the household level, not the person level. Selected demographic characteristics such as age, educational attainment, and ethnicity are reported for the head of the household. However, the individual considered to be the head of the household may not be the same individual who filled out and returned the questionnaire.<sup>24</sup> To the extent that this is the case, there would be less reason to expect a strong relationship between the demographic information provided for the head of the household and responses provided by the (possibly different) individual who filled out the questionnaire. The second is that household-level commercial data are known to be of limited accuracy (cf. Roth, Han, and Montaquila 2013). Even if, in every case, the head of the household was the person who responded to the questionnaire, substantial inaccuracies in the auxiliary variables could further undermine any associations that might otherwise exist (West 2013).

Because of the finding in chapter 2 that households missing data for any auxiliary variable were significantly underrepresented in the respondent pool, section 3.2 evaluated the potential efficacy of targeting households based on the extent to which they are missing auxiliary data. The analysis found that households' propensity to respond tends to decrease as the number of key variables for which auxiliary data are missing increases. It also found statistically significant relationships between the number of auxiliary variables for which auxiliary data were missing for a household and that household's propensity to report a certification/license or current enrollment in college classes. The direction of the relationships was consistent with the direction of the estimated bias for these items—specifically, households with more missing auxiliary variables were less likely to report certifications/licenses, and more likely to report college enrollment. This suggests that if substantial improvements in unit response rates among households with missing data for multiple auxiliary variables could be achieved, unit nonresponse bias in these particular estimates may be reduced. A relationship between missing auxiliary data and certificate reporting was also observed, but only among NRFU respondents (i.e. NATES nonrespondents).

On balance, therefore, **these results provide some support, albeit limited, for targeting incentives and/or nonresponse follow-up at households for which the sampling frame is missing data for multiple key auxiliary variables.** A major comparative advantage of a strategy that targets households

---

<sup>24</sup> Equivalently, for households in which multiple eligible individuals returned the NATES questionnaire, the individual treated as the head of the household in the auxiliary variables may not be the same as the individual whose questionnaire was retained for the purpose of this analysis.

based on the fact that a certain number of auxiliary variables are missing, rather than on the value of those variables, is that its efficacy would not be dependent on the accuracy of the commercially provided data. Additional research is needed to determine whether higher incentives and/or more intensive nonresponse follow-up procedures can succeed at increasing response rates among these households, and whether the observed associations between the number of missing auxiliary variables and reporting of key estimates are consistent across administrations. It should also be noted that, for certificates, the direction of the relationship among NRFU respondents suggests some risk that increased response rates among cases missing auxiliary data could *increase* the observed bias in the prevalence estimate. As discussed in section 2.2 of chapter 2, a negative (though not statistically significant) bias was observed in the proportion of households reporting certificates. Since NATES nonrespondents with two or more missing auxiliary variables appear to be less likely to report certificates than NATES nonrespondents with fewer than two missing auxiliary variables, converting a substantial number of such cases to NATES respondents could, in principle, reduce the NATES estimate of certificate prevalence and thus exacerbate the observed bias. This result illustrates the difficulty of developing a comprehensive approach to bias reduction in a survey with multiple key estimates, as relationships with available auxiliary data are likely to vary between estimates.

*This page is intentionally left blank.*

# Chapter 4: Analysis of Item Nonresponse Bias

Item nonresponse occurs when a sample member returns a questionnaire but fails to respond to a particular item to which he or she was eligible to respond. This chapter reports the results of a two-stage analysis of potential item nonresponse bias that was conducted using data from respondents to the mailed NATES questionnaire. These two stages are reported in sections 4.1 and 4.2, respectively. Section 4.3 provides a discussion of the results.

Because missing rates for the key survey items ranged from 1.3 to 2.8 percent, there was little risk of item nonresponse bias. However, missing rates were higher, in many cases substantially more so, for items related to the detailed characteristics of reported credentials. Because these items were skip-dependent, the question of interest in this chapter is whether respondents' difficulty in correctly following the skip patterns on the self-administered NATES questionnaire may have increased the risk of item nonresponse bias. While complex skip patterns are less likely to affect item response in telephone surveys (where interviewers and computer-assisted telephone interviewing technology can help respondents navigate the questionnaire), they are likely to be more of a problem on a self-administered questionnaire. Indeed, subsequent cognitive testing by the American Institutes for Research (2013) has suggested that some respondents have difficulty following skip patterns similar to those included in NATES.

The following criteria were used to select items for this analysis:

- The items were located in the core substantive sections of the NATES questionnaire (that is, the sections pertaining to high school diplomas and college degrees, certifications/licenses, educational certificates, apprenticeships, college classes, and other work-related trainings);
- The items had weighted missing rates above 15 percent, the threshold suggested by the *NCES Statistical Standards* (Seastrom 2012); and
- The items were located immediately following a stem item (that is, an item that directed some respondents to skip subsequent items based on their response).

Items immediately following stem items were chosen for the analysis of item nonresponse bias because their response rates were likely to have been particularly affected by respondent confusion over skip patterns.

The following 12 items meeting the three above criteria were selected for analysis:

- Activities to earn continuing education units for most recent certification/license (Q12)
- Certification/license was for past or future job (Q18)<sup>25</sup>
- Certificate was a subbaccalaureate certificate (Q21)
- Year certificate was earned (Q22)

---

<sup>25</sup> Q18 consisted of two separate “yes/no” items: Q18a (credential was obtained for a past job) and Q18b (credential was obtained for a future job). Cases were classified as nonrespondents to Q18 only if they left both component items blank.

- Certificate was for past or future job (Q28)<sup>26</sup>
- Year of apprenticeship (Q30)
- Apprenticeship occupation (Q34)
- College classes are to earn a diploma, certificate, or degree (Q36)
- Diploma, certificate, or degree being earned (Q37)
- Number of classes currently taking (Q39)
- Employer reimbursing tuition (Q44)
- Training provided at no charge (Q50)

Item nonresponse analysis was also performed on the English ability item (Q78). Although this item was not in one of the core substantive sections of the NATES questionnaire, its estimates are reported in table 2.3 of this report, and the *NCES Statistical Standards* (Seastrom 2012) require that item nonresponse analysis be performed on any item with a missing rate above 15 percent for which estimates are published.<sup>27</sup>

It is important to note that the analysis in this chapter does not permit the direct measurement of nonresponse bias in the estimate of interest. However, it does permit an evaluation of whether item respondents are different from nonrespondents in ways that would be expected to affect the distribution of reported responses to the item. If, for any given item, characteristics found in the first stage (identifying differences between item respondents and nonrespondents) to be predictive of item response status are also found in the second stage (assessing the risk of item nonresponse bias) to be predictive of the actual response to the item, then item nonresponse bias is likely to exist. Under the assumption that similar associations between demographic characteristics and the response to the item would have existed among item nonrespondents had they answered the item, these two analyses can provide an indication of the *risk* of item nonresponse bias in the NATES estimates.

## 4.1 Differences Between Item Respondents and Nonrespondents

In this section, differences between item respondents and nonrespondents are assessed in terms of the following characteristics, as reported on the questionnaire: educational attainment, whether the respondent speaks a language other than English at home, age, and employment status. Educational attainment, household language (as a proxy for English ability), and age are all hypothesized to be related to the degree of difficulty that a respondent may have had in following the skip patterns on the questionnaire. Employment status is included as a proxy for topic salience. Because NATES measured the prevalence of work-related credentials, respondents who were not employed at the time they filled out the questionnaire may have viewed the topic as being less relevant to them and read the questionnaire less

<sup>26</sup> As with Q18, cases were classified as nonrespondents to Q28 only if they left both component items blank.

<sup>27</sup> In a full-scale administration, an item nonresponse bias analysis would be conducted for every item with a missing rate above 15 percent. Because NATES was a pilot study conducted with the purpose of refining the survey questionnaire, the item nonresponse bias analysis focused on items for which missing rates were hypothesized to be related to a key feature of the questionnaire design, namely, the skip patterns. The assumption was made that the branch items following the items analyzed in this report would show similar patterns of item nonresponse bias. In addition, a full-scale administration would typically include an analysis to determine whether imputation reduced item nonresponse bias. This step was not feasible for the NATES pilot study because no items were imputed. Altogether, there were 25 NATES items for which weighted missing rates exceeded 15 percent and item nonresponse bias analysis was not performed.



carefully. Therefore, these respondents may be expected to be more likely to accidentally skip an item. For each item, the percentage distribution of each of the four independent variables is estimated separately for item respondents and item nonrespondents. The statistical significance of the overall association between each demographic characteristic and item response status is evaluated using a design-adjusted test of independence.<sup>28</sup>

The results of this stage of the analysis are summarized in table 4.1. For each item analyzed, table 4.1 lists the characteristics for which the test of independence showed a statistically significant relationship ( $p < .05$ ) between the characteristic and the item response status. The table also provides overall weighted missing rates for each item as a percentage of NATES respondents who, based on their previous response to the stem item, were eligible to answer the item. Tables with complete results for each item analyzed are provided in appendix B of this report.

**Table 4.1. Item missing rates and characteristics associated with item response status, by demographic characteristic and item: 2013**

Item	Item missing rate (percent)	Associated with response status			
		Education <sup>1</sup>	Household language <sup>2</sup>	Age <sup>3</sup>	Employment status <sup>4</sup>
Q12 (activities to earn CEUs)	25.0	X		X	X
Q18a and Q18b (certification/license for past/future job)	40.1	X	X		X
Q21 (subbaccalaureate certificate)	24.6	X			
Q22 (year of certificate)	27.8	X			
Q28a and Q28b (certificate for past/future job)	39.7	X	X		X
Q30 (year of apprenticeship)	37.7	X	X		
Q34 (apprenticeship occupation)	63.4	X			
Q36 (classes to earn credential)	21.6	X		X	
Q37 (credential being earned)	23.9	X		X	
Q39 (number of classes)	78.1	X			
Q44 (employer reimbursing tuition)	84.2	X			X
Q50 (no charge for training)	17.9	X		X	X
Q78 (English ability)	15.5	X	†		X

† Not applicable (household language item was the stem item for Q78).

<sup>1</sup> The respondent's highest reported level of education (Q1): high school diploma, high school equivalent, or lower, some college or Associate's degree, or Bachelor's degree or higher.

<sup>2</sup> Whether the respondent speaks a language other than English at home (Q77).

<sup>3</sup> The respondent's reported age (Q71): 16–34, 35–44, 45–54, or 55–65.

<sup>4</sup> Whether the respondent is currently employed (Q52).

NOTE: Item missing rates represent the weighted percent of NATES respondents who did not answer the item, with valid skips excluded.

An "X" indicates that a statistically significant relationship existed between item response status and the specified characteristic.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

<sup>28</sup> In an unweighted analysis with an assumed simple random sample, the Pearson test of independence calculates a chi-squared statistic. Hence, it is commonly known as the "chi-squared test of independence." However, the statistical software package used for this analysis corrects for the complex NATES sampling design by converting the chi-squared statistic into an  $F$  statistic. The interpretation is the same as that of the chi-squared test: namely, a significant  $p$  value implies that the two categorical variables are not independent.

For all items, a statistically significant difference exists between item respondents and nonrespondents in terms of the educational attainment distribution. Substantially higher proportions of nonrespondents than respondents reported holding a high school diploma or lower, while substantially lower proportions of nonrespondents reported holding a bachelor's degree or higher. In other words, there is a positive relationship between educational attainment and item response propensity. For most items, differences between respondents and nonrespondents in the prevalence of the "high school or lower" and/or "bachelor's degree or higher" subgroups are in the range of 10 to 40 percentage points.

Significant differences in terms of the household language exist for three items, with nonrespondents more likely by roughly 10 percentage points to report speaking a language other than English at home. Significant differences in terms of age exist for four items. For all of these items, nonrespondents are older than respondents, with the "55 to 65" proportion being higher among nonrespondents by 10 to 25 percentage points.

In contrast, while six items show a statistically significant difference between respondents and nonrespondents in terms of employment status, only four follow the expected pattern, with the employed proportion lower among nonrespondents than among respondents. This implies that topic salience does not affect item response propensity in as consistent a manner as do educational attainment, age, and language.

Thus, each of these key demographic characteristics is associated with item response status for at least some of the 13 items analyzed. In general, participants who are less educated, who are older, and/or who speak a language other than English at home appear to be less likely to answer items that immediately follow skips. This is consistent with the hypothesis that these respondents may have had more difficulty following the skip patterns in the NATES questionnaire. While the direction of the association between employment status and item response status is less consistent, an association does exist for approximately half of the items.

## 4.2 Analysis of Item Nonresponse Bias

The second stage of the item nonresponse analysis uses design-adjusted tests of independence to evaluate the relationship between the same four demographic characteristics and the distribution of the reported response to each item. For each of the 10 items analyzed in this stage,<sup>29</sup> table 4.2 identifies the characteristics for which the following two conditions hold:

- The first stage of the analysis found a statistically significant (.05 level or below) difference in the distribution of the characteristic between item respondents and nonrespondents; and
- The second stage found a statistically significant (.05 level or below) association between the characteristic and the response chosen by respondents to the item.

Again, tables with detailed results for each item are provided in appendix B. The summary results in table 4.2 suggest that differences in educational attainment and, to a lesser extent, employment status are the

---

<sup>29</sup> Because Q34 was an uncoded write-in item, analysis of the distribution of responses was not feasible. In addition, because of the high missing rates on Q39 and Q44, the sample sizes were insufficient to meet NCES reporting standards. Thus, Q39 and Q44 were also excluded from this stage of the analysis. For Q37 and Q78, response categories were collapsed in order to meet NCES reporting standards.

most likely drivers of item nonresponse bias in the NATES estimates. Education is associated with both response status and the distribution of responses for seven items. The same is true for age for one item and for employment status for three items.

Altogether, of the 10 items analyzed in this stage, there are eight for which at least one demographic characteristic is predictive both of response status and of the actual response to the item. This implies that respondents and nonrespondents to the items differ from each other in ways that would be expected to affect the distribution of responses to the items—that is, to generate item nonresponse bias.

For the other two items (Q22 and Q28), the first stage did find that significant differences between item respondents and nonrespondents existed in terms of educational attainment, household language, and/or employment status. Insofar as there is no corresponding effect of these characteristics on the actual responses to the items, the second stage provides no evidence that these differences led to item nonresponse bias.

**Table 4.2. Characteristics associated with response status and response to item, by demographic characteristics and item: 2013**

Item	Associated with response status and response to item			
	Education <sup>1</sup>	Household language <sup>2</sup>	Age <sup>3</sup>	Employment status <sup>4</sup>
Q12 (activities to earn CEUs)	X			X
Q18a and Q18b (certification/license for past/future job)	X			
Q21 (subbaccalaureate certificate)	X			
Q22 (year of certificate)				
Q28a and Q28b (certificate for past/future job)				
Q30 (year of apprenticeship)	X			
Q36 (classes to earn credential)	X		X	
Q37 (credential being earned)	X			
Q50 (no charge for training)				X
Q78 (English ability)	X	†		X

† Not applicable (household language item was the stem item for Q78).

<sup>1</sup> The respondent's highest reported level of education (Q1): high school diploma, high school equivalent, or lower, some college or Associate's degree, or Bachelor's degree or higher.

<sup>2</sup> Whether the respondent speaks a language other than English at home (Q77).

<sup>3</sup> The respondent's reported age (Q71): 16–34, 35–44, 45–54, or 55–65.

<sup>4</sup> Whether the respondent is currently employed (Q52).

NOTE: An "X" indicates that a statistically significant association existed between item response status and the specified characteristic *and* between the response to the item and the specified characteristic.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

## 4.3 Discussion of Results

The first stage of the analysis, reported in section 4.1, found that significant demographic differences existed between item respondents and nonrespondents for all 13 of the items analyzed. In particular, item nonrespondents tended to be less educated, to be older, and/or to speak a language other than English at home. Employment status was also found to be associated with nonresponse to several items, but in a less consistent direction.

Because the analysis focused on low-response items that immediately followed stem items, the results are consistent with the hypothesis that participants within certain educational, age, and language subgroups may have had difficulty following the complex skip logic in the NATES questionnaire. On their own, however, the results do not imply the presence of item nonresponse bias, because they provide no indication of the likelihood that item nonrespondents would have provided a different distribution of responses had they answered the items.

Consequently, a follow-up analysis, reported in section 4.2, evaluated the extent to which education, age, household language, and employment status were also predictive of actual responses to the items. For eight of the 10 items analyzed at this stage, at least one of the characteristics that had been previously found to be associated with response status was also found to be associated with the response to the item. It is important to emphasize that this stage of the analysis could be conducted only for item respondents, and it is impossible to determine whether the same relationships between demographic characteristics and the response to the item would have been observed among item nonrespondents. Thus, it still cannot be stated with certainty that these eight items were affected by nonresponse bias (or, conversely, that the other two were not). Nevertheless, on the basis of this two-stage analysis, it can be stated that the available information is consistent with a risk of nonresponse bias for a large proportion of the items analyzed. The fact that these demographic variables are related both to item response status and to the response to the item suggests that, in a full-scale administration, imputation models that included them as independent variables would be expected to reduce item nonresponse bias.

This analysis focused on a subset of items for which high missing rates were hypothesized to be related to respondent confusion over the NATES skip logic. The results reported in this chapter therefore **provide support for minimizing skip patterns in future mailed administrations of ATEs and other household surveys**. More generally, the results provide evidence that item nonresponse in self-administered surveys is not always random, but rather may be driven by characteristics that are also predictive of the actual response to the item. Consequently, in order to obtain valid estimates from mailed questionnaires, the minimization of item nonresponse should, as a general rule, be a key driver of major questionnaire design decisions. This includes decisions related to the number and complexity of skips, but also includes decisions as to the length of the questionnaire, the reading level of the text, the formatting of the questionnaire form, and other factors that could plausibly be expected to affect item response rates.

## Chapter 5: Conclusion

The purpose of this report was to use data from the 2013 National Adult Training and Education Survey (NATES) to evaluate the extent of unit and item nonresponse bias in a mailed household survey focused on adults' non-degree credentials and work-related education and training. The original NATES questionnaire achieved a household response rate of 65.0 percent and a final person-level response rate of 62.5 percent. However, in-person follow-up interviews with a subsample of nonrespondents, referred to as the Nonresponse Follow-up Study (NRFU) sample, achieved a response rate of 83.1 percent. The use of a shortened version of the NATES questionnaire for these follow-up interviews enabled the direct measurement of unit nonresponse bias in key NATES estimates. Taking advantage of this uniquely rich set of data on nonrespondents, chapter 2 of this report addressed the following research questions focused on unit nonresponse bias:

- **Question 1: Measuring unit nonresponse bias before nonresponse adjustment.** To what extent are key survey estimates derived from selection-weighted NATES respondents affected by unit nonresponse bias?
- **Question 2: Assessing the efficacy of nonresponse adjustment.** Is unit nonresponse bias substantially reduced when the sampling weights for NATES respondents are adjusted for unit nonresponse?
- **Question 3: Comparing alternative methods of unit nonresponse bias analysis.** Do methods that are available for nonresponse bias analysis in the absence of an NRFU provide an accurate indication of the risk of bias and of the effectiveness of nonresponse adjustment?

Chapter 3 evaluated two potential strategies for reducing unit nonresponse bias in future administrations of this survey:

- **Question 4: Improving weighting adjustments for nonresponse.** Is there reason to expect that the effectiveness of the weighting adjustments at reducing nonresponse bias could be improved if more complete data were obtained for the demographic characteristics available in or linked to the sampling frame?
- **Question 5: Increasing response rates among key subgroups.** Is there reason to expect that nonresponse bias would be reduced if certain subgroups of sampled households were chosen to receive larger incentives, more intensive nonresponse follow-up, or other targeted efforts to increase response rates?

For a selection of NATES items with low response rates, chapter 4 of this report also addressed the following research questions related to item nonresponse bias:

- **Question 6: Identifying differences between item respondents and nonrespondents.** Were any demographic characteristics consistently associated with nonresponse to skip-dependent items?

- **Question 7: Assessing the risk of item nonresponse bias.** For each of the items evaluated, is there reason to expect that differences between item respondents and nonrespondents led to changes in the distribution of reported responses to the item?

The remainder of this chapter summarizes key findings related to each of these research questions while also highlighting important limitations of this study. On the basis of these findings, it then offers recommendations for future ATEs administrations.

## 5.1 Summary of Findings

### ***Question 1: Measuring unit nonresponse bias before nonresponse adjustment***

Two distinct methods were used to measure unit nonresponse bias. First, households that returned one or more NATES questionnaires were compared on a number of demographic characteristics to the entire national sample of eligible households from the NATES sampling frame. While this method cannot directly estimate unit nonresponse bias for NATES estimates, it did provide an indication of which demographic subgroups were over- and underrepresented in the sample of respondents, relative to their share of the eligible population. When the proportions for respondent households were estimated using unadjusted selection weights, large differences between respondent households and the eligible sample were found in terms of nearly every variable of interest. For example, households in low-minority or low-poverty strata, households with street addresses, and households with high incomes, were overrepresented in the respondent pool. The heads of respondent households were significantly more likely to be male, to be White, to have a bachelor's or graduate degree, and to be married. Respondent households were also less likely to have missing auxiliary data for any given variable.

Second, respondents to the mailed NATES questionnaire were compared to the combined sample of respondents to the mailed NATES questionnaire and the NRFU follow-up interviews in terms of their responses to key survey items included in both questionnaires. Because the combined selection-weighted sample of NATES and NRFU respondents was roughly representative of the entire eligible population, proportions generated from this sample were treated as the “true” population estimates for the purpose of measuring nonresponse bias. When the proportions for NATES respondents were estimated using selection weights, unit nonresponse bias was found to be not statistically significant for the certification/license and certificate attainment items. However, the proportion currently enrolled in college-level classes was underestimated by about 1.5 percentage points, which was statistically significant at the .05 level. In general, estimates of bias were higher and statistically significant for demographic characteristics such as education, race, age, and income.

### ***Question 2: Assessing the efficacy of nonresponse adjustment***

The household-level nonresponse adjustments to the NATES weights appeared to be more effective at reducing bias in auxiliary variables than at reducing bias in the survey variables. When the distributions of auxiliary variables were re-estimated using nonresponse-adjusted weights for households that responded to the mailed NATES questionnaire, the estimated nonresponse bias decreased by an average of 40.7 percent relative to the selection-weighted proportions. With that said, for all but a few variables,

statistically significant bias in excess of one percentage point was still observed in the estimated proportions of some subcategories.

When the distributions of responses to key survey items were re-estimated for NATES respondents using household-level nonresponse-adjusted weights, the estimated nonresponse bias decreased by an average of 19.9 percent. The downward bias in the estimated proportion of the population currently enrolled in college-level classes decreased to about 0.9 percentage points and became not statistically significant. In other words, after nonresponse adjustment, there was no statistically significant bias in any of the three key survey items included on both questionnaires: certifications/licenses, certificates, and college enrollment.

Bias in the estimated proportions of several educational, age, racial, and income subgroups decreased slightly, but remained statistically significant. In other words, for the demographic estimates that showed the most bias prior to adjustment, the household-level nonresponse adjustments had little practical effect. However, it should be noted that the remaining bias in demographic proportions could be mitigated through the use of a poststratification or raking adjustment, as high-quality demographic estimates are available for the NATES target population from sources such as the decennial census and the American Community Survey (ACS). These results therefore suggest that nonignorable bias is likely to be limited in the NATES proportions that were evaluated using the NRFU.

For NATES items not included in the NRFU questionnaire, it was not possible to directly calculate unit nonresponse bias. However, for several of these items, the selection-weighted and nonresponse-adjusted percentage distributions of responses were compared in order to provide further insight into the efficacy of the nonresponse adjustment process. The adjustments led to changes in estimates that were not statistically significant in all but one of the items evaluated. Assuming that these items exhibit nonresponse bias comparable in magnitude to that observed in the other credential attainment and demographic items, this suggests that any reduction in bias from the person-level nonresponse adjustments was minimal. However, given the fact that no statistically significant bias was found in the selection-weighted certification/license and certificate proportions, it is possible that there was also little or no bias to be removed in the selection-weighted apprenticeship and training proportions.

### ***Question 3: Comparing alternative methods of unit nonresponse bias analysis***

The NRFU was a one-time feature of the NATES pilot study and is not planned to be repeated in future, full-scale ATEs administrations. In the absence of an NRFU, methods that are available for nonresponse bias analysis include those discussed in sections 2.1 (comparing the distribution of auxiliary variables available in or linked to the sampling frame) and 2.3 (comparing selection-weighted and nonresponse-adjusted key estimates). The results of this study found that, relative to the use of an NRFU, the use of auxiliary variables for nonresponse bias analysis appears to overestimate the risk of bias in key NATES estimates. This is consistent with the finding (discussed in chapter 3) that the available auxiliary variables show little association with credential reporting, and with prior research (discussed in section 1.3 of chapter 1) that has found substantial measurement error in commercially purchased demographic variables. Additionally, because many of the auxiliary variables used for bias analysis were also used to define the nonresponse adjustment cells, the weighting adjustments led to large changes in bias as measured by auxiliary variables, but relatively small changes in bias measured using the NRFU. The results also indicate that a simple comparison of selection-weighted and nonresponse-adjusted key

estimates is unlikely to provide an accurate measure of nonresponse bias. For the proportions for which bias could be measured using the NRFU, the nonresponse adjustments removed a relatively small fraction of the observed bias. Consequently, the change in an estimate attributable to nonresponse adjustment appears likely to substantially underestimate the magnitude of the bias in the selection-weighted estimate.

The results of this study therefore reinforce the findings of Groves and Peytcheva (2008) that estimates of nonresponse bias tend to be sensitive to the analytical method used. While the use of an NRFU is not without its limitations (as discussed in section 5.2), it is the only available method that entails the collection of data for key survey items from a representative or near-representative sample of nonrespondents. If the use of an NRFU is therefore treated as the “gold standard” for nonresponse bias analysis, these results suggest that the two other methods used in this study may not provide accurate measures of the risk of nonresponse bias in key survey estimates. Additional research is needed to evaluate whether other potential methods—for example, using later respondents as a proxy for nonrespondents—are able to better approximate the results obtained using an NRFU.

#### ***Question 4: Improving weighting adjustments for nonresponse***

Two possible strategies for addressing the possibility of unit nonresponse bias in future administrations were considered. The first would be to increase the proportion of the sample for which auxiliary data on key demographic characteristics are available. This strategy would rely on the assumption that the relative ineffectiveness of the nonresponse adjustment process is attributable to the high missing data rates for available auxiliary variables. In order to test this assumption, it was necessary to determine whether, for households for which data on these variables were available, these variables were strongly associated with responses to key NATES items. A logistic regression analysis found that, among the auxiliary variables that were used in the NATES nonresponse adjustment process, the age of the head of the household was significantly associated with the propensity to report college class enrollment or work-related training. No other auxiliary variables available in the NATES sampling frame were consistently associated with the propensity to report more than one credential or type of training.

Altogether, these results suggest that the low effectiveness of the nonresponse adjustment process is driven not only by missing data for the auxiliary variables, but also by low underlying correlations between these variables and key survey estimates. Thus, augmenting auxiliary data for these variables is unlikely to substantially improve the effectiveness of the nonresponse adjustments. It is also worth noting that, although this study focused on a statistical evaluation of this strategy rather than on an evaluation of its operational feasibility, commercial sources of household-level demographic data are, in general, quite limited. Therefore, it is unclear whether a reliable and cost effective source of supplemental data could be found for households that were missing data in the original sampling frame. West, Wagner, Hubbard, and Gu (2013) compared auxiliary data from two separate commercial vendors. Because both sources showed high missing rates, combining the two sources (for variables available from both) partially but not fully mitigated the missing data problem. Additionally, for cases with data available from both sources, there was a high rate of disagreement between the sources, which is consistent with the findings of prior studies that commercial demographic data are prone to measurement error.



### ***Question 5: Increasing response rates among key subgroups***

A second strategy to reduce nonresponse bias would be to attempt to increase unit response rates through the use of incentives, nonresponse follow-up, and other methods. Because such methods are already used in NCES household surveys, a particularly relevant question is whether they could be targeted to specific low-response subgroups. In order to reduce nonresponse bias, response rates would need to be increased specifically among subgroups that are likely to provide a different distribution of responses than existing respondents. The weak association between auxiliary variables and key survey estimates, combined with the fact that the auxiliary variables are of unknown accuracy, suggests that targeting households based on the demographic characteristics available in the sampling frame is likely to be of limited effectiveness.

However, it is also possible to target households based on whether substantial amounts of auxiliary data are missing for the household. A comparison of unit response rates by the number of missing auxiliary variables confirmed that the more auxiliary variables a household was missing, the less likely the household was to return a questionnaire (households with more missing auxiliary data also showed higher address ineligibility rates). Again, however, the efficacy of this strategy would depend on whether households that are missing large amounts of auxiliary data are also likely to provide different responses to key items. A logistic regression analysis found no relationship between the number of missing auxiliary variables and the propensity to report a apprenticeship or work-related training. However, households missing two or more auxiliary variables were significantly less likely to report a certification/license and significantly more likely to report enrollment in college classes. This suggests that an improvement in response rates among these households could have reduced nonresponse bias in those estimates. Missing auxiliary data was also related to the propensity to report a certificate, but only among NRFU respondents (e.g., NATES nonrespondents).

### ***Question 6: Identifying differences between item respondents and nonrespondents***

For 13 items from the mailed NATES questionnaire, item respondents were compared to item nonrespondents in terms of the distribution of four reported demographic characteristics: educational attainment, age, whether a language other than English is spoken in the household, and employment status. The selected items were all items with high missing rates that immediately followed a skip directive. These items were selected in order to evaluate whether confusion over skip patterns may have led to consistent patterns of item nonresponse.

Educational attainment was significantly related to response status for every item, with substantially higher proportions of nonrespondents than respondents reporting holding a high school diploma or lower. Age and household language were significantly related to response status for four and three items, respectively, and followed the expected pattern: larger proportions of nonrespondents reported being over the age of 55 or speaking a language other than English at home. Employment status was also associated with item response for slightly less than half of the items, although in less consistent directions. Consistent with expectations, these findings demonstrate that NATES respondents who were less educated, who were older, and/or who spoke a language other than English at home were less likely to answer items immediately following skips.

### **Question 7: Assessing the risk of item nonresponse bias**

In order to determine whether demographic differences between item respondents and nonrespondents may have led to item nonresponse bias, the relationship between the same four demographic characteristics and the response selected by item respondents was evaluated. For eight items, one or more characteristics previously found to be related to nonresponse were also found to be significantly associated with the choice of responses. This implies that respondents and nonrespondents to the items differ from each other in ways that would be expected to affect the distribution of responses. This finding is therefore consistent with a risk of item nonresponse bias in many NATES items that immediately followed skip directives. It also suggests, however, that imputation models that use these four demographic variables as independent variables would be expected to reduce nonresponse bias in these items.

## **5.2 Limitations**

Several potential limitations of this study should be noted.

### ***Nonresponse to the follow-up interviews of NATES nonrespondents***

The analysis of unit nonresponse bias in the NATES estimates relied on the assumption that the combined respondents to the mailed NATES questionnaires and the in-person NRFU interviews were representative of the eligible population; and, therefore, that proportions estimated using the NATES plus NRFU respondents were reasonable approximations of the true population proportions. However, because the NRFU follow-up interviews did not achieve a 100 percent response rate, they are unlikely to have been perfectly representative of the population of NATES nonrespondents. To the extent that the NRFU proportions themselves are affected by nonresponse bias, measures of nonresponse bias in NATES constructed using the follow-up sample will be inaccurate.

A limited analysis of nonresponse bias in the NRFU sample itself, using auxiliary variables, is provided in appendix A of this report. The results indicate that significant differences existed between NRFU respondents and nonrespondents in terms of the ethnicity and poverty strata, and the ethnicity and age of the head of the household as reported in the sampling frame. However, because nonresponse bias is a function of both the response rate and differences between respondents and nonrespondents, the high response rate (83.1 percent) to the NRFU follow-up interviews suggests that the risk of substantial nonresponse bias in the NRFU data is low. It should be noted, however, that the response rate to the NRFU was below the 85 percent threshold at which the *NCES Statistical Standards* (Seastrom 2012) would require a nonresponse bias analysis in a full-scale administration.

### ***Other types of nonsampling bias***

It is also possible that the NATES and/or NRFU proportions were affected by other types of nonsampling bias, such as measurement error and coverage bias. For any given proportion, the measures of unit nonresponse bias presented in this report attribute the entire difference between the NATES proportion and the NATES plus NRFU proportion to unit nonresponse bias. This method does not require the assumption that other types of nonsampling bias are nonexistent. However, it does require the assumption that, for any given proportion, bias in the NATES proportion is of the same direction and magnitude as

bias in the NATES plus NRFU proportion. If this is not the case, then the measures of unit nonresponse bias presented in this report may be inaccurate, because part of the difference between the proportions may actually be attributable to forms of bias other than unit nonresponse bias.

### ***Mode effects***

An additional limitation of the use of the NRFU data to estimate unit nonresponse bias is that it relies on the assumptions that responses to key items were not subject to mode effects. In other words, it assumes that every respondent to an in-person NRFU interview gave the same response to every item that would have been given had he or she filled out a paper questionnaire. However, to the extent that responses to key items were affected by the interview mode, differences between the NATES-only and NATES plus NRFU proportions would not be entirely attributable to unit nonresponse bias. The presence of an interviewer can, in theory, affect survey outcomes through a number of channels, such as social desirability bias and the ability of the interviewer to provide additional information (cf. de Leeuw 2002). It should be noted, however, that the field interviewers were not provided with any probes other than the information provided in the question text, which was the same text included on the mailed NATES questionnaire.

### ***Within-household selection of NATES respondents***

Although some households returned multiple NATES questionnaires, only one questionnaire per household was retained in any analysis that made use of data from the NRFU follow-up sample. This includes the measurement of nonresponse bias in chapter 2 and the regressions in chapter 3. The NATES data were collapsed in this way to enable comparability with the NRFU data, which were collected from only one (non-randomly selected) person in each household. Additional detail about this approach and its rationale is provided in section E.5 of appendix E.

It is important to note that because of this analytical approach, the proportions shown in chapter 2 cannot be interpreted as person-level estimates for the NATES target population. As discussed in section E.5 of appendix E, to minimize differences in selection bias between the NATES and NRFU samples and thus avoid confounding estimates of nonresponse bias, the basis for the comparison was the “first, most convenient” reporter in each household. Due to the non-random nature of within-household selection, the proportions shown in chapter 2 should be interpreted as household-level estimates of the characteristics of the “first, most convenient” reporter.

In contrast to this approach, future full-scale administrations of ATES will use a two-phase sample with a household roster to allow for random within-household selection of eligible persons and thus permit valid person-level estimation. For this reason, estimates of credential prevalence shown in chapter 2 may not be directly comparable to estimates obtained from future ATES administrations, or from other nationally representative surveys that include similar items. It is also important to note that, because no household roster was collected, this approach implicitly relies on the assumption that the person who filled out the “first” mailed questionnaire would have been the person to respond to the in-person interview had the household been sampled for the NRFU, which cannot be empirically verified. To the extent that this assumption is inaccurate, estimates of nonresponse bias may still be confounded by differences in selection bias.

### ***Inability to directly measure item nonresponse bias***

For the analysis of item nonresponse bias in chapter 4, the relationship between demographic characteristics and the distribution of responses could only be estimated among item respondents. Thus, the conclusion that item nonresponse bias is likely to exist in a particular item relies on the assumption that relationships that exist among respondents also exist among nonrespondents. Because there was no means of determining how item nonrespondents would have responded to any of the items analyzed, the validity of this assumption could not be tested, nor could item nonresponse bias be estimated directly.

## **5.3 Summary of Recommendations**

This section discusses four recommendations that follow from the results presented in this report:

- **Recommendation 1:** Target households with missing auxiliary data for multiple variables
- **Recommendation 2:** Use screener data for nonresponse adjustments and/or response rate targeting
- **Recommendation 3:** Explore alternative sources of auxiliary data on respondents and nonrespondents
- **Recommendation 4:** Minimize item nonresponse bias through simplified questionnaire design

### ***Recommendation 1: Target households with missing auxiliary data for multiple variables***

Household-level response rates to NATES tended to decrease as the number of auxiliary variables for which households were missing data increased. Furthermore, the number of missing auxiliary variables was found to be associated with certification/license attainment and college enrollment in a manner consistent with the direction of the measured nonresponse bias in those items. These findings suggest that nonresponse bias in these items would be reduced if response rates among such households were more comparable to response rates among households with more complete auxiliary data.

Consequently, on the basis of this study, **the number of missing auxiliary variables may be usable as a criterion for the targeting of strategies such as increased incentives or more intensive nonresponse follow-up. Additional research should be conducted to determine whether such strategies lead to increased response rates among these households, and whether the association between missing auxiliary data and credential reporting remains consistent across ATEs administrations.** A caveat to this conclusion is that if addresses with substantial missing auxiliary data are more likely to turn out to be ineligible—as was the case in NATES—intensive targeting of these households could potentially entail a waste of resources, which could offset the benefits of increased response rates. Additionally, even if such a strategy helped to minimize bias in the certification/license and college enrollment estimates, there is no guarantee that this would be true for all estimates of interest. For example, the relationship between missing auxiliary data and certificate reporting was found to be more complex, and may suggest that targeting cases missing auxiliary data would raise a risk of increasing bias in the certificate prevalence estimate. In general, because relationships between available auxiliary data and key survey variables are likely to be variable-specific, any design in which data collection strategies are targeted based on auxiliary data must be informed by careful consideration of the relevant relationships and the relative priority assigned to particular estimates.

As discussed by Harter et al. (2016), the inability to match commercial data to an address tends to be associated with neighborhood-level characteristics such as lower incomes, larger immigrant populations, or indicators of high population mobility. As shown in appendix C of this report, these relationships are observed in the NATES sample—NATES-sampled households with missing data for two or more auxiliary variables were, on average, located in Census tracts with higher proportions of rented housing units, higher proportions of non-English speakers, and lower median household incomes, among other differences. Regression analysis suggests that a high proportion of residents who have recently moved and a high prevalence of multi-unit housing structures are the strongest independent predictors of missingness in the auxiliary data. Therefore, the area-level characteristics for which missing data appears to be a proxy may also provide a useful criterion for targeting data collection efforts.

### ***Recommendation 2: Use screener data for nonresponse adjustments and/or response rate targeting***

This study suggests that the demographic variables available in commercially purchased address-based sampling frames are of limited use, either for generating effective nonresponse adjustments or for targeting response rate maximization techniques. This limited utility results not only from high missing data rates, but also from weak correlations between these variables and responses to key survey items.

Despite the limitations of commercial auxiliary data, few alternative means exist of obtaining household-level demographic data for both respondents and nonrespondents. However, future ATES administrations will differ from NATES in that they will use a two-stage sampling design. Sampled households will first receive a short screener form on which they will be asked to report basic demographic data (primarily focused on age and school enrollment status) on every household member. For each household that returns a screener form, one eligible member will be selected to receive the ATES topical questionnaire. This is in contrast to the NATES sampling design, in which every sampled household was sent multiple topical questionnaires with no screener stage.

While the use of screener data for nonresponse adjustment was not empirically evaluated in this report, this recommendation is rooted in the apparent disutility of the commercial auxiliary data. Information reported on the screener may be more useful than commercial auxiliary data for reducing nonresponse bias on the topical questionnaire. First, self-reported data are likely to be more accurate than commercially provided data. Second, the screener will provide data on every household member, not just the head of the household. Information reported on the screener for the sampled household member may therefore correlate more strongly with that individual's responses on the topical questionnaire.

For this reason, in NCES administrations that use a screener stage, information from the screener form is typically used to generate person-level nonresponse adjustments. **Consideration should also be given to using the information reported on the screener form to target incentives or follow-up at the topical stage of data collection.**

A limitation of this approach is that screener forms typically attempt to minimize respondent burden by including only a small set of items about each household member. **It may be possible to improve the efficacy of the screener data at reducing nonresponse bias by expanding the form to include items expected to correlate strongly with responses to key items at the topical stage. However, this**

**consideration must be balanced with the need to minimize response burden** at the screener stage, as previous research has shown that longer screeners result in lower response rates.

Another inherent limitation of screener data is that they are only available for households that return screeners. Consequently, while they can be used in the person-level weighting adjustments and to target incentives or follow-up at the topical stage, they cannot be used in the generation of household-level nonresponse-adjusted weights or for targeting at the screener stage.

### ***Recommendation 3: Explore alternative sources of auxiliary data on respondents and nonrespondents***

The results in this study suggest that the efficacy of the nonresponse adjustments could not be substantially improved by obtaining data on the auxiliary variables for a larger proportion of cases. However, this study did not explore the potential efficacy of obtaining data on a wider set of *variables*. For example, data for a rich set of demographic variables from the decennial census and the American Community Survey are publicly available at the block group or tract level. If the sampling frame contained block group and/or tract identifiers (as the NATES sampling frame did), or if addresses were geocoded to block groups or tracts, these data could be appended to the sampling frame and provide information on a wider set of characteristics for respondents and nonrespondents.

A likely obstacle to this approach is that, insofar as household-level demographic characteristics correlate weakly with responses by individuals to survey items, data at a higher level of aggregation would intuitively be expected to show even less of an association. Consistent with this expectation, Biemer and Peytchev (2013) found that geocoded census data did not generate effective nonresponse adjustments for the National Comorbidity Survey Replication (a study sponsored by the National Institute of Mental Health).

However, because of the relatively low cost of obtaining and appending census data, **additional research should be conducted to assess their potential efficacy at correcting nonresponse bias** in future ATES administrations.

### ***Recommendation 4: Minimize item nonresponse bias through simplified questionnaire design***

A precondition for the collection of high-quality data using self-administered questionnaires is the avoidance of questionnaire design features that can lead to item nonresponse. This study focused specifically on the possible extent of item nonresponse bias in items that immediately followed skip directives. The findings are consistent with the hypothesis that many respondents, particularly those with lower education levels, had difficulty navigating the skip patterns in the NATES questionnaire. They also show that characteristics predictive of item nonresponse were often associated with the responses chosen by item respondents. The implication is that the NATES skip patterns may have contributed to item nonresponse bias.

Therefore, **the results of this study support the minimization of complex skip patterns in self-administered questionnaires**. In response to the high missing rates for some NATES items, as well as findings from cognitive interviews, subsequent versions of the ATES questionnaire have been substantially simplified.

# References

- Amaya, A., LeClere, F., Fiorio, L., and English, N. (2014). Improving the Utility of the DSF Address-Based Frame through Ancillary Information. *Field Methods*, 26(1): 70–86.
- American Association for Public Opinion Research. (2015). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Deerfield, IL: AAPOR.
- American Institutes for Research. (2013). *Cognitive Interview Findings for Items to Measure Certifications, Licenses, and Educational Certificates Among Adults in the United States*. Report prepared for the National Center for Education Statistics. Retrieved April 15, 2015 from [https://nces.ed.gov/surveys/gemena/pdf/FINAL\\_ATES\\_cog\\_int\\_report\\_Dec\\_18\\_2013.pdf](https://nces.ed.gov/surveys/gemena/pdf/FINAL_ATES_cog_int_report_Dec_18_2013.pdf).
- Biemer, P., and Peytchev, A. (2013). Using Geocoded Census Data for Nonresponse Bias Correction: An Assessment. *Journal of Survey Statistics and Methodology*, 1(1): 24–44.
- Bielick, S., Cronen, S., Stone, C., Montaquila, J.M., and Roth, S.B. (2013). *The Adult Training and Education Survey (ATES) Pilot Study Technical Report* (NCES 2013-190). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- de Leeuw, E.D. (2002). *Data Quality in Mail, Telephone, and Face-to-Face Surveys*. Amsterdam: TT-Publikaties.
- Deming, W.E. (1953). On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Nonresponse and the Bias of Nonresponse. *Journal of the American Statistical Association*, 48(264): 743–772.
- DiSogra, C., Dennis, J.M., and Fahimi, M. (2010). *On the Quality of Ancillary Data Available for Address-Based Sampling*. Paper presented at the 2010 Joint Statistical Meetings. Retrieved October 19, 2016 from <http://www.knowledgenetworks.com/ganp/docs/jsm2010/On-the-Quality-of-Ancillary-ABS-2010-JSM-submission.pdf>.
- Federal Committee on Statistical Methodology. (2001). *Measuring and Reporting Sources of Error in Surveys* (Statistical Policy Working Paper 31). Washington, DC: Office of Management and Budget, Statistical Policy Office.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72(2): 167–189.
- Hansen, M.H. and Hurwitz, W.N. (1946). The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association*, 41(236): 517–529.
- Harter, R., Battaglia, M.P., Buskirk, T.D., Dillman, D.A., English, N., Fahimi, M., Frankel, M.R., Kennel, T., McMichael, J.P., McPhee, C.B., Montaquila, J., Yancey, T., and Zukerberg, A.L. (2016).

- Address-Based Sampling*. Oakbrook Terrace, IL: American Association for Public Opinion Research.
- Harter, R.M., Mach, T.L., Chapline, J.F., and Wolken, J.D. (2007). *Determining Subsampling Rates for Nonrespondents*. Paper presented at the Third International Conference on Establishment Surveys. Retrieved November 4, 2016 from <http://ww2.amstat.org/meetings/ices/2007/proceedings/ICES2007-000197.PDF>.
- Heeringa, S.G., West, B.T., and Berglund, P.A. (2010). *Applied Survey Data Analysis*. Boca Raton, FL: CRC Press.
- Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- McPhee, C., Bielick, S., Masterton, M., Flores, L., Parmer, R., Amchin, S., Stern, S., and McGowan, H. (2015). *National Household Education Surveys Program of 2012: Data File User's Manual* (NCES 2015-030). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- National Institutes of Health, National Institute of Mental Health. (2001 – 2003). *National Comorbidity Survey Replication (NCS-R)*. Data available at <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/20240>.
- Pasek, J., Jang, S.M., Cobb, C.L., Dennis, J.M., and DiSogra, C. (2014). Can Marketing Data Aid Survey Research? Examining Accuracy and Completeness in Consumer-File Data. *Public Opinion Quarterly*, 78(4): 889–916.
- Roth, S.B., Han, D., and Montaquila, J.M. (2013). The ABS Frame: Quality and Considerations. *Survey Practice*, 6(4).
- Roth, S.B., Montaquila, J.M., and Chapman, C. (2006). *Nonresponse Bias in the 2005 National Household Education Surveys Program* (NCES 2007-016). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Rubin, D.B. (1976). Inference and Missing Data. *Biometrika*, 63(3): 581–592.
- SAS Institute. (2011). *SAS/STAT 9.3 User's Guide*. Cary, NC: SAS Institute.
- Seastrom, M. (2012). *NCES Statistical Standards* (NCES 2014-097). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Seastrom, M., Kaufman, S., and Lee, R. (2012). Appendix A: Imputation. In *NCES Statistical Standards* (NCES 2014-097). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Tersine, A. and Starsinic, M. (2003). *Optimum Nonresponse Subsampling Rate for the American Community Survey*. Paper presented at the 2003 Joint Statistical Meetings. Retrieved November 4, 2016 from [https://www.census.gov/content/dam/Census/library/working-papers/2003/acs/2003\\_Tersine\\_01\\_wpd.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2003/acs/2003_Tersine_01_wpd.pdf).



- U.S. Department of Commerce, Bureau of the Census. (2005 – present). *American Community Survey (ACS)*. Data available at <https://www.census.gov/programs-surveys/acs/data.html>.
- U.S. Department of Commerce, Bureau of the Census. (2013). *Wave 13 of the 2008 Survey of Income and Program Participation (SIPP)*. Data available at <http://www.census.gov/programs-surveys/sipp/data/2008-panel/wave-13.html>.
- U.S. Department of Commerce, Bureau of the Census. (2015). *Current Population Survey (CPS)*. Data available at [http://thedataweb.rm.census.gov/ftp/cps\\_ftp.html](http://thedataweb.rm.census.gov/ftp/cps_ftp.html).
- U.S. Department of Education, National Center for Education Statistics. (1988). *National Educational Longitudinal Study (NELS)*. Data available at [https://nces.ed.gov/surveys/nels88/data\\_products.asp](https://nces.ed.gov/surveys/nels88/data_products.asp).
- U.S. Department of Education, National Center for Education Statistics. (2010). *Adult Training and Education Survey (ATES) Pilot Study*. [no publicly available dataset]
- U.S. Department of Education, National Center for Education Statistics. (2012). *National Household Education Survey (NHES)*. Data available at <http://nces.ed.gov/nhes/dataproducts.asp>.
- U.S. Department of Education, National Center for Education Statistics. (2013). *National Adult Training and Education Survey (NATES)*. [no publicly available dataset]
- U.S. Department of Education, National Center for Education Statistics. (2013). *National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU)*. [no publicly available dataset]
- U.S. Department of Education, National Center for Education Statistics. (2014). *National Household Education Survey (NHES) Feasibility Study*. [no publicly available dataset]
- U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (2007). *National Intimate Partner and Sexual Violence Survey (NISVS) Pilot Study*. [no publicly available dataset]
- U.S. Department of Justice, Bureau of Justice Statistics. (2013). *National Crime Victimization Survey (NCVS)*. Data available at <http://www.bjs.gov/developer/ncvs/index.cfm>.
- Van de Kerckhove, W., Krenzke, T., and Mohadjer, L. (2009). *Adult Literacy and Lifeskills Survey (ALL) 2003: U.S. Nonresponse Bias Analysis* (NCES 2009-063). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- West, B.T., Wagner, J., Hubbard, F., and Gu, H. (2015). The Utility of Alternative Commercial Data Sources for Survey Operations and Estimation: Evidence from the National Survey of Family Growth. *Journal of Survey Statistics and Methodology*, 3(2): 240–264.
- West, B.T. (2013). An Examination of the Quality and Utility of Interviewer Observations in the National Survey of Family Growth. *Journal of the Royal Statistical Society Series A*, 176(1), 211–225.
- Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

*This page is intentionally left blank.*

# Appendix A: Nonresponse Bias in the Nonresponse Follow-up Sample

The Nonresponse Follow-up Study (NRFU), which consisted of follow-up in-person interviews with a subsample of National Adult Training and Education Survey (NATES) nonrespondent households, achieved a response rate of 83.1 percent. While this relatively high response rate reduces the risk of nonresponse bias, it does not eliminate it. In this report, NRFU respondents were assumed to be representative of the population of NATES nonrespondents; consequently, selection-weighted estimates generated using combined NATES and NRFU respondents were assumed to be reasonable estimates of the true population values. To the extent that nonresponse bias is present in the NRFU itself, this assumption may be undermined.

This appendix reports the results of a basic analysis of unit nonresponse bias in the NRFU follow-up study. Because no follow-up was conducted with nonrespondents to the NRFU, the sampling frame is the only source of data for both respondents and nonrespondents. Consequently, the methodology used in this appendix is the same as that described in section 2.1 of chapter 2. Respondents to the NRFU are compared to all eligible sampled households in terms of the distribution of auxiliary variables available in or linked to the sampling frame. Note that, in this analysis, the eligible sample refers to households that were sampled *for the NRFU*, which was a subsample of the nonrespondents to the mailed NATES questionnaire.

This method identifies the types of households that are underrepresented in the respondent pool relative to their share of the eligible sample. Under the assumption that households that differ in terms of these characteristics would also have differed in terms of their answers to the NRFU questionnaire items, this analysis can provide an indication of the risk of unit nonresponse bias in the NRFU.

Table A.1 compares the percentage distribution of the NRFU eligible sample to the percentage distribution of NRFU respondents in terms of 14 variables available in the sampling frame. For each proportion, the estimated bias (column 3) is defined as the difference between the respondents-only proportion (column 2) and the eligible sample proportion (column 1). The percent relative bias (column 4) is defined as the ratio of the estimated bias to the eligible sample proportion. The table also reports a *t* test for the statistical significance of the estimated bias (column 5).

Relative to their share of the eligible sample, households from the low-minority and low-poverty strata are significantly underrepresented in the pool of NRFU respondents. Households for which age data on the head of the household are missing, and households for which the head of the household is Black or Hispanic, are significantly overrepresented. No other proportion reported in table A.1 shows statistically significant bias. In contrast, in the comparable analysis of the mailed NATES survey reported in section 2.1, statistically significant differences between respondents and the eligible sample exist in terms of nearly every auxiliary variable analyzed.

These results suggest that the households that responded to the NRFU were largely representative of the sampled households in terms of key demographic characteristics. This, combined with the relatively high response rate to the NRFU, suggests that NRFU respondents were, as intended, largely representative of the population of NATES nonrespondents.

**Table A.1. Percentage distribution of NRFU respondents and eligible sample, estimated bias, and percent relative bias, by selected auxiliary variables: 2013**

Selected auxiliary variable	Percent of eligible sample 1	Percent of respondents 2	Estimated bias 3	Percent relative bias 4	t test of bias 5
Race/ethnicity stratum					
Black	19.6	20.3	0.7	3.6	1.2
Hispanic	16.1	16.8	0.7	4.2	1.4
Other	64.2	62.9	-1.4	-2.2	-2.1*
Ethnicity of the head of household					
Missing	42.1	41.8	-0.4	-0.9	-0.6
White	34.0	33.1	-0.9	-2.6	-1.2
Black	7.2	8.0	0.8	10.9	3.2*
Hispanic	10.4	11.3	0.9	9.0	2.6*
Asian	3.1	2.5	-0.6	-19.4	-1.7
Other	3.2	3.4	0.1	3.9	0.6
Household income					
Missing	21.3	21.4	0.1	0.5	0.2
\$0–\$10,000	3.7	3.3	-0.4	-10.0	-1.2
\$10,001–\$20,000	6.4	6.4	0.0	0.4	0.1
\$20,001–\$30,000	8.0	8.4	0.4	5.0	1.2
\$30,001–\$40,000	9.8	9.9	0.1	0.8	0.2
\$40,001–\$50,000	7.3	7.7	0.4	5.5	1.3
\$50,001–\$60,000	9.0	9.0	-0.1	-0.7	-0.1
\$60,001–\$75,000	9.0	9.0	-0.1	-0.7	-0.2
\$75,001–\$100,000	10.9	11.0	0.1	0.8	0.2
\$100,001–\$150,000	8.8	8.7	-0.1	-1.2	-0.2
\$150,001+	5.6	5.1	-0.5	-8.8	-1.3
Education of the head of household					
Missing	45.9	46.1	0.2	0.4	0.3
Less than high school diploma	12.3	13.0	0.7	6.0	1.8
High school diploma	11.2	11.3	0.1	1.2	0.3
Some college	16.6	17.2	0.6	3.7	1.3
Bachelor's degree	9.2	8.3	-0.9	-9.6	-1.6
Graduate degree	4.9	4.1	-0.8	-16.5	-1.5

See notes at end of table.

**Table A.1. Percentage distribution of NRFU respondents and eligible sample, estimated bias, and percent relative bias, by selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	Percent of eligible sample 1	Percent of respondents 2	Estimated bias 3	Percent relative bias 4	t test of bias 5
Age of the head of household					
Missing	50.4	51.8	1.4	2.9	2.0*
18–24	1.8	1.5	-0.3	-14.7	-1.2
25–34	6.5	6.8	0.3	4.5	1.1
35–44	11.5	11.5	0.1	0.7	0.2
45–54	14.4	13.5	-0.9	-6.4	-1.5
55–65	10.5	9.6	-0.8	-8.1	-1.6
Over 65	5.0	5.2	0.2	4.3	0.6
Census tract poverty rate					
High	23.2	24.5	1.3	5.7	3.0*
Not high	76.8	75.5	-1.3	-1.7	-3.0*
Phone number available in sampling frame					
Yes	32.2	32.1	-0.1	-0.3	-0.1
No	67.8	67.9	0.1	0.1	0.1
Route type					
High rise	32.8	32.7	-0.1	-0.2	-0.1
Street	67.2	67.3	0.1	0.1	0.1
Number of adults in household					
Missing	21.6	21.8	0.2	0.7	0.3
1	39.6	39.4	-0.2	-0.5	-0.3
2	21.6	22.3	0.8	3.6	1.5
3	10.5	9.9	-0.6	-5.4	-1.3
4	4.2	3.8	-0.3	-7.9	-1.0
5	1.7	1.8	0.1	5.6	0.8
6	‡	‡	‡	‡	‡
7	‡	‡	‡	‡	‡
8	‡	‡	‡	‡	‡
Home tenure					
Missing	27.3	27.8	0.5	1.9	0.8
Own	47.2	45.8	-1.4	-2.9	-1.7
Rent	25.5	26.4	0.9	3.4	1.6
Census region					
Northeast	20.9	20.8	-0.0	-0.1	0.0
South	38.3	38.8	0.6	1.5	0.9
Midwest	16.2	16.0	-0.2	-1.1	-0.3
West	24.7	24.4	-0.4	-1.5	-0.5

See notes at end of table.

**Table A.1. Percentage distribution of NRFU respondents and eligible sample, estimated bias, and percent relative bias, by selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	Percent of eligible sample 1	Percent of respondents 2	Estimated bias 3	Percent relative bias 4	t test of bias 5
Gender of the head of household					
Missing	26.0	26.2	0.2	0.7	0.3
Female	27.1	27.8	0.7	2.5	1.0
Male	46.9	46.0	-0.9	-1.9	-1.1
Marital status of the head of household					
Missing	34.2	34.6	0.4	1.1	0.6
Married	37.4	37.7	0.3	0.9	0.4
Single	28.5	27.7	-0.7	-2.5	-1.0
Dwelling type					
Single-family unit	62.3	62.2	-0.2	-0.3	-0.2
Multi-unit	37.7	37.8	0.2	0.5	0.2

‡ Reporting standards not met. Either there are too few cases for a reliable estimate or the coefficient of variation (CV) is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of eligible sampled households or respondent households with the specified characteristic. "Estimated bias" represents the difference between the respondent proportion and the eligible sample proportion. "Percent relative bias" represents the estimated bias divided by the eligible sample proportion. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

# Appendix B: Detailed Tables From Item Nonresponse Analysis

This appendix contains detailed tables for both stages of the item nonresponse analysis conducted for the National Adult Training and Education Survey (NATES) and discussed in chapter 4. The appendix is divided into two sections. Section B.1 reports the distribution of key demographic characteristics among respondents and nonrespondents to each of the 13 items analyzed in the first stage. Section B.2 reports the distribution of responses to each of the 10 items analyzed in the second stage, broken down by the same key demographic characteristics. Refer to chapter 4 for a summary of the results and implications of this analysis.

## B.1 Differences Between Item Respondents and Item Nonrespondents

**Table B.1. Percentage distribution of respondents and nonrespondents to Q12 (activities to earn continuing education units), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	9.2	33.5	6.2*
Some college or associate's degree	28.7	30.2	0.4
Bachelor's degree or higher	62.1	36.4	-7.7*
Pearson F statistic	40.3*		
Speaks a language other than English at home			
Yes	14.9	19.7	1.5
No	85.1	80.3	-1.5
Pearson F statistic	2.6		
Age			
16–34	24.4	18.2	-1.6
35–44	25.3	14.9	-3.4*
45–54	26.1	31.0	1.0
55–65	24.3	35.9	3.6*
Pearson F statistic	5.1*		
Employment status			
Employed	86.1	70.9	-3.1*
Not employed	13.9	29.1	3.1*
Pearson F statistic	12.9*		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson F test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.2. Percentage distribution of respondents and nonrespondents to Q18a/Q18b (certification/license for past or future job), by selected reported characteristics: 2013**

<b>Selected reported characteristic</b>	<b>Percent of item respondents</b>	<b>Percent of item nonrespondents</b>	<b><i>t</i> statistic</b>
Highest educational attainment			
High school, high school equivalent, or lower	12.8	35.8	5.1*
Some college or associate's degree	44.4	34.5	-2.2*
Bachelor's degree or higher	42.8	29.7	-3.3*
Pearson <i>F</i> statistic	17.9*		
Speaks a language other than English at home			
Yes	14.0	22.9	2.6*
No	86.1	77.1	-2.6*
Pearson <i>F</i> statistic	8.2*		
Age			
16–34	20.1	18.6	-0.4
35–44	20.3	17.5	-0.8
45–54	24.3	26.3	0.5
55–65	35.4	37.7	0.5
Pearson <i>F</i> statistic	0.3		
Employment status			
Employed	46.5	60.0	2.6*
Not employed	53.5	40.0	-2.6*
Pearson <i>F</i> statistic	6.8*		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.



**Table B.3. Percentage distribution of respondents and nonrespondents to Q21 (subbaccalaureate certificate), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	<i>t</i> statistic
Highest educational attainment			
High school, high school equivalent, or lower	23.2	47.7	5.1*
Some college or associate's degree	56.5	40.9	-3.2*
Bachelor's degree or higher	20.4	11.4	-3.5*
Pearson <i>F</i> statistic	19.8*		
Speaks a language other than English at home			
Yes	20.6	24.0	1.0
No	79.4	76.1	-1.0
Pearson <i>F</i> statistic	1.0		
Age			
16–34	22.0	15.1	-1.9
35–44	16.8	15.5	-0.4
45–54	31.7	31.9	0.1
55–65	29.5	37.5	1.6
Pearson <i>F</i> statistic	1.6		
Employment status			
Employed	69.0	63.9	-1.0
Not employed	31.0	36.1	1.0
Pearson <i>F</i> statistic	1.1		

\* $p < .05$

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.4. Percentage distribution of respondents and nonrespondents to Q22 (year certificate was earned), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	25.8	46.8	4.2*
Some college or associate's degree	59.8	41.5	-3.6*
Bachelor's degree or higher	14.4	11.7	-1.1
Pearson F statistic	12.3*		
Speaks a language other than English at home			
Yes	18.8	23.2	1.3
No	81.2	76.8	-1.3
Pearson F statistic	1.7		
Age			
16–34	22.2	13.8	-2.3*
35–44	16.5	15.8	-0.2
45–54	30.9	32.6	0.3
55–65	30.4	37.9	1.4
Pearson F statistic	1.7		
Employment status			
Employed	67.8	64.8	-0.6
Not employed	32.2	35.2	0.6
Pearson F statistic	0.4		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson F test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.5. Percentage distribution of respondents and nonrespondents to Q28a/Q28b (certificate for past or future job), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	27.0	43.7	3.1*
Some college or associate's degree	60.2	44.1	-2.9*
Bachelor's degree or higher	12.8	12.2	-0.2
Pearson <i>F</i> statistic	6.4*		
Speaks a language other than English at home			
Yes	16.5	25.3	2.3*
No	83.5	74.7	-2.3*
Pearson <i>F</i> statistic	5.4*		
Age			
16–34	20.4	17.6	-0.7
35–44	16.1	14.6	-0.4
45–54	30.9	31.1	0.0
55–65	32.7	36.7	0.7
Pearson <i>F</i> statistic	0.3		
Employment status			
Employed	48.5	65.1	2.7*
Not employed	51.5	34.9	-2.7*
Pearson <i>F</i> statistic	7.0*		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.6. Percentage distribution of respondents and nonrespondents to Q30 (year of last apprenticeship), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	<i>t</i> statistic
Highest educational attainment			
High school, high school equivalent, or lower	23.2	52.1	5.4*
Some college or associate's degree	47.6	28.1	-4.0*
Bachelor's degree or higher	29.2	19.9	-2.2*
Pearson <i>F</i> statistic	15.1*		
Speaks a language other than English at home			
Yes	17.2	27.2	2.5*
No	82.8	72.8	-2.5*
Pearson <i>F</i> statistic	7.6*		
Age			
16–34	23.3	17.3	-1.3
35–44	18.0	15.1	-0.7
45–54	28.4	32.5	0.5
55–65	30.3	35.1	0.8
Pearson <i>F</i> statistic	0.7		
Employment status			
Employed	67.4	69.3	0.3
Not employed	32.6	30.7	-0.3
Pearson <i>F</i> statistic	0.1		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.7. Percentage distribution of respondents and nonrespondents to Q34 (apprenticeship occupation), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	17.2	54.0	5.6*
Some college or associate's degree	55.5	26.9	-4.2*
Bachelor's degree or higher	27.3	19.1	-1.3
Pearson <i>F</i> statistic	13.9*		
Speaks a language other than English at home			
Yes	18.4	26.2	1.5
No	81.6	73.8	-1.5
Pearson <i>F</i> statistic	2.2		
Age			
16–34	23.1	14.7	-1.5
35–44	12.3	15.0	0.6
45–54	22.9	37.9	1.7
55–65	41.8	32.5	-1.4
Pearson <i>F</i> statistic	2.3		
Employment status			
Employed	44.2	60.7	1.7
Not employed	55.8	39.4	-1.7
Pearson <i>F</i> statistic	3.0		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.8. Percentage distribution of respondents and nonrespondents to Q36 (classes to earn credential), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	12.0	64.5	9.8*
Some college or associate's degree	60.9	22.2	-8.2*
Bachelor's degree or higher	27.1	13.3	-4.2*
Pearson F statistic	75.2*		
Speaks a language other than English at home			
Yes	25.9	24.8	-0.2
No	74.1	75.2	0.2
Pearson F statistic	0.0		
Age			
16–34	73.6	30.9	-8.2*
35–44	14.7	12.5	-0.5
45–54	6.6	29.9	3.6*
55–65	5.2	26.8	4.0*
Pearson F statistic	30.0*		
Employment status			
Employed	63.7	69.5	0.8
Not employed	36.3	30.5	-0.8
Pearson F statistic	0.7		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson F test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.9. Percentage distribution of respondents and nonrespondents to Q37 (credential being earned), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	11.9	62.8	9.3*
Some college or associate's degree	63.2	24.3	-8.0*
Bachelor's degree or higher	25.0	12.8	-3.5*
Pearson F statistic	63.7*		
Speaks a language other than English at home			
Yes	25.4	25.9	0.1
No	74.6	74.1	-0.1
Pearson F statistic	0.0		
Age			
16–34	76.9	33.5	-8.1*
35–44	14.3	12.0	-0.5
45–54	5.3	28.0	3.6*
55–65	3.5	26.4	4.2*
Pearson F statistic	32.3*		
Employment status			
Employed	64.0	68.1	0.6
Not employed	36.0	32.2	-0.6
Pearson F statistic	0.3		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson F test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.10. Percentage distribution of respondents and nonrespondents to Q39 (number of classes), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	12.7 !	63.5	7.9*
Some college or associate's degree	33.9	22.5	-1.3
Bachelor's degree or higher	53.4	14.1	-4.1*
Pearson F statistic	23.2*		
Speaks a language other than English at home			
Yes	30.6	24.1	-0.7
No	69.4	75.9	0.7
Pearson F statistic	0.6		
Age			
16–34	32.7	31.6	-0.1
35–44	‡	‡	‡
45–54	23.2	29.7	0.8
55–65	23.4	26.4	0.4
Pearson F statistic	0.7		
Employment status			
Employed	63.3	69.2	0.6
Not employed	36.7	30.9	-0.6
Pearson F statistic	0.4		

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

‡ Reporting standards not met. Either there were too few cases for a reliable estimate or the coefficient of variation (CV) is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson F test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.



**Table B.11. Percentage distribution of respondents and nonrespondents to Q44 (employer reimbursing tuition), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	‡	60.8	2.0
Some college or associate's degree	24.3 !	24.0	0.0
Bachelor's degree or higher	47.3	15.2	-2.3*
Pearson F statistic	4.6*		
Speaks a language other than English at home			
Yes	27.7 !	24.9	-0.2
No	72.3	75.1	0.2
Pearson F statistic	0.1		
Age			
16–34	32.3	32.9	0.1
35–44	‡	‡	‡
45–54	32.2 !	26.2	-0.5
55–65	18.1 !	27.4	1.1
Pearson F statistic	0.4		
Employment status			
Employed	94.9	70.6	-3.5*
Not employed	‡	29.4	3.5*
Pearson F statistic	8.7*		

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

‡ Reporting standards not met. Either there were too few cases for a reliable estimate or the coefficient of variation (CV) is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic.

The Pearson F test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.12. Percentage distribution of respondents and nonrespondents to Q50 (training provided at no charge), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	<i>t</i> statistic
Highest educational attainment			
High school, high school equivalent, or lower	19.0	39.6	4.7*
Some college or associate's degree	32.5	34.0	0.3
Bachelor's degree or higher	48.5	26.5	-5.7*
Pearson <i>F</i> statistic	18.1*		
Speaks a language other than English at home			
Yes	17.4	21.9	1.3
No	82.2	78.1	-1.3
Pearson <i>F</i> statistic	1.7		
Age			
16–34	30.9	28.1	-0.7
35–44	24.7	10.8	-4.0*
45–54	25.2	29.7	1.0
55–65	19.3	31.5	2.7*
Pearson <i>F</i> statistic	6.1*		
Employment status			
Employed	91.1	76.8	-3.3*
Not employed	8.9	23.3	3.3*
Pearson <i>F</i> statistic	18.3*		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.13. Percentage distribution of respondents and nonrespondents to Q78 (English ability), by selected reported characteristics: 2013**

Selected reported characteristic	Percent of item respondents	Percent of item nonrespondents	t statistic
Highest educational attainment			
High school, high school equivalent, or lower	35.7	48.7	2.7*
Some college or associate's degree	29.1	23.3	-1.8
Bachelor's degree or higher	35.2	28.0	-1.7
Pearson <i>F</i> statistic	4.8*		
Age			
16–34	34.1	26.6	-1.4
35–44	21.3	19.7	-0.4
45–54	25.0	23.7	-0.3
55–65	19.6	30.1	2.4*
Pearson <i>F</i> statistic	2.4		
Employment status			
Employed	68.9	50.2	-4.5*
Not employed	31.1	49.8	4.5*
Pearson <i>F</i> statistic	22.1*		

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents or nonrespondents with the specified characteristic. The Pearson *F* test determines whether a statistically significant relationship exists between response status and education, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

## B.2 Response Distributions by Demographic Subgroups

**Table B.14. Percentage of NATES respondents reporting specified responses to Q12 (activities to earn continuing education units), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting				
	Not yet had to meet requirements	Conference	Class/seminar	Instructional materials	Other
Highest educational attainment					
High school, high school equivalent, or lower	14.9 !	14.3	50.3	17.7 !	‡
Some college or associate's degree	8.3	20.0	52.4	14.1	5.2 !
Bachelor's degree or higher	9.0	35.2	44.3	7.8	3.8
Pearson <i>F</i> statistic	3.9*				
Speaks a language other than English at home					
Yes	13.3	26.2	47.8	10.3	2.6 !
No	8.6	29.5	47.2	10.4	4.3
Pearson <i>F</i> statistic	1.0				
Age					
16–34	15.8	23.5	42.5	11.7	6.5 !
35–44	6.6	34.0	48.9	7.8	‡
45–54	8.7	32.2	45.2	10.7	3.2
55–65	6.8	25.9	52.0	11.5	4.0
Pearson <i>F</i> statistic	1.8				
Employment status					
Employed	8.6	31.3	46.3	10.2	3.6
Not employed	13.8 !	13.7	52.5	13.0	7.1 !
Pearson <i>F</i> statistic	4.0*				

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

‡ Reporting standards not met. Either there were too few cases for a reliable estimate or the coefficient of variation (CV) is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.15. Percentage of NATES respondents reporting specified responses to Q18a/b (certification/license for past/future job), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting			
	Past job	Future job	Both past and future job	Neither past nor future job
Highest educational attainment				
High school, high school equivalent, or lower	35.3	9.2 !	28.5	27.1 !
Some college or associate's degree	54.1	16.5	22.4	7.0
Bachelor's degree or higher	58.8	9.7	22.0	9.5
Pearson <i>F</i> statistic	3.0*			
Speaks a language other than English at home				
Yes	39.7	16.6	29.7	14.0 !
No	55.8	12.3	21.8	10.1
Pearson <i>F</i> statistic	1.9			
Age				
16–34	31.1	31.8	24.1	13.1
35–44	44.8	8.8 !	40.6	5.8 !
45–54	46.7	6.8 !	28.0	18.5 !
55–65	74.3	9.2 !	9.0	7.6 !
Pearson <i>F</i> statistic	6.3*			
Employment status				
Employed	52.6	14.2	22.1	11.2
Not employed	54.6	11.9	23.8	9.7 !
Pearson <i>F</i> statistic	0.2			

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.16. Percentage of NATES respondents reporting specified responses to Q21 (type of certificate), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting	
	Post-baccalaureate certificate	Sub-baccalaureate certificate
Highest educational attainment		
High school, high school equivalent, or lower	6.5 !	93.5
Some college or associate's degree	9.0	91.0
Bachelor's degree or higher	39.1	60.9
Pearson <i>F</i> statistic	34.9*	
Speaks a language other than English at home		
Yes	22.5	77.5
No	12.3	87.7
Pearson <i>F</i> statistic	8.5*	
Age		
16–34	14.9	85.1
35–44	15.7	84.3
45–54	16.0	84.0
55–65	11.6	88.4
Pearson <i>F</i> statistic	0.5	
Employment status		
Employed	15.6	84.4
Not employed	11.5	88.5
Pearson <i>F</i> statistic	1.5	

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status.

Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.17. Percentage of NATES respondents reporting specified responses to Q22 (year of last certificate), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting	
	Before 2009	2009–2013
Highest educational attainment		
High school, high school equivalent, or lower	70.5	29.5
Some college or associate's degree	74.8	25.2
Bachelor's degree or higher	77.7	22.3
Pearson <i>F</i> statistic	0.5	
Speaks a language other than English at home		
Yes	71.7	28.3
No	74.6	25.4
Pearson <i>F</i> statistic	0.3	
Age		
16–34	44.2	55.8
35–44	77.1	22.9
45–54	81.2	18.8
55–65	86.8	13.2
Pearson <i>F</i> statistic	23.8*	
Employment status		
Employed	76.0	24.0
Not employed	69.8	30.2
Pearson <i>F</i> statistic	1.1	

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.18. Percentage of NATES respondents reporting specified responses to Q28a/b (certificate for past or future job), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting			
	Past job	Future job	Both past and future job	Neither past nor future job
Highest educational attainment				
High school, high school equivalent, or lower	55.4	12.2	10.4	22.0
Some college or associate's degree	56.9	18.9	10.8	13.4
Bachelor's degree or higher	58.1	15.2 !	11.4 !	‡
Pearson <i>F</i> statistic	0.5			
Speaks a language other than English at home				
Yes	36.7	21.8 !	14.8	26.7 !
No	60.3	15.5	10.1	14.2
Pearson <i>F</i> statistic	2.6			
Age				
16–34	19.9 !	39.2	20.6	20.3
35–44	35.4	18.2	19.1	27.3
45–54	65.8	9.5 !	5.7 !	19.1
55–65	79.9	8.4 !	5.6 !	6.1 !
Pearson <i>F</i> statistic	7.9*			
Employment status				
Employed	56.9	13.2	10.3	19.7
Not employed	57.3	18.9	10.8	12.9
Pearson <i>F</i> statistic	0.9			

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

‡ Reporting standards not met. Either there were too few cases for a reliable estimate or the coefficient of variation (CV) is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.



**Table B.19. Percentage of NATES respondents reporting specified responses to Q30 (year of most recent apprenticeship), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting	
	Before 2009	2009–2013
Highest educational attainment		
High school, high school equivalent, or lower	84.5	15.6
Some college or associate's degree	87.3	12.7
Bachelor's degree or higher	65.8	34.2
Pearson <i>F</i> statistic	6.8*	
Speaks a language other than English at home		
Yes	74.0	26.0
No	81.2	18.8
Pearson <i>F</i> statistic	1.0	
Age		
16–34	45.8	54.2
35–44	86.2	13.8 !
45–54	86.6	13.4 !
55–65	96.8	3.2 !
Pearson <i>F</i> statistic	20.5*	
Employment status		
Employed	76.8	23.2
Not employed	86.9	13.1
Pearson <i>F</i> statistic	3.1	

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.20. Percentage of NATES respondents reporting specified responses to Q36 (classes to earn credential), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting	
	Yes	No
Highest educational attainment		
High school, high school equivalent, or lower	90.5	9.5 !
Some college or associate's degree	95.3	4.7
Bachelor's degree or higher	83.7	16.4
Pearson <i>F</i> statistic	7.9*	
Speaks a language other than English at home		
Yes	90.5	9.5
No	91.6	8.4
Pearson <i>F</i> statistic	0.2	
Age		
16–34	96.2	3.8
35–44	88.8	11.2
45–54	71.1	28.9
55–65	64.1	35.9
Pearson <i>F</i> statistic	25.9*	
Employment status		
Employed	91.9	8.2
Not employed	91.6	8.4
Pearson <i>F</i> statistic	0.0	

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded.

Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.21. Percentage of NATES respondents reporting specified responses to Q37 (credential earned), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting		
	Below a bachelor's degree	Bachelor's degree	Above a bachelor's degree
Highest educational attainment			
High school, high school equivalent, or lower	64.2	32.3	‡
Some college or associate's degree	41.3	55.8	2.9 !
Bachelor's degree or higher	10.1	15.4	74.5
Pearson <i>F</i> statistic	75.7*		
Speaks a language other than English at home			
Yes	26.9	53.7	19.4
No	39.8	38.4	21.8
Pearson <i>F</i> statistic	4.1*		
Age			
16–34	34.7	45.4	20.0
35–44	38.4	40.9	20.8
45–54	40.9	27.4	31.7
55–65	‡	‡	‡
Pearson <i>F</i> statistic	2.2		
Employment status			
Employed	31.4	43.7	25.0
Not employed	45.2	41.2	13.6
Pearson <i>F</i> statistic	4.6*		

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

‡ Reporting standards not met. Either there were too few cases for a reliable estimate or the coefficient of variation (CV) is 50 percent or greater.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.22. Percentage of NATES respondents reporting specified responses to Q50 (training provided by employer), by response and selected reported characteristics: 2013**

Selected reported characteristic	Percent of NATES respondents reporting		
	Yes	No	Not relevant
Highest educational attainment			
High school, high school equivalent, or lower	83.0	11.6	5.5 !
Some college or associate's degree	78.5	13.3	8.2
Bachelor's degree or higher	78.0	16.7	5.3
Pearson <i>F</i> statistic	1.3		
Speaks a language other than English at home			
Yes	79.5	12.4	8.1
No	79.2	15.2	5.6
Pearson <i>F</i> statistic	1.4		
Age			
16–34	79.3	15.9	4.8
35–44	84.1	11.2	4.7 !
45–54	78.8	13.5	7.7
55–65	74.4	17.7	7.9
Pearson <i>F</i> statistic	1.2		
Employment status			
Employed	82.7	13.6	3.7
Not employed	44.2	24.6 !	31.2
Pearson <i>F</i> statistic	33.0*		

! Interpret data with caution. The coefficient of variation (CV) for this estimate is between 30 and 50 percent.

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table B.23. Percentage of NATES respondents reporting specified responses to Q78 (how well English spoken), by response and selected reported characteristics**

Selected reported characteristic	Percent of NATES respondents reporting	
	Very well or well	Not very well or not well at all
Highest educational attainment		
High school, high school equivalent, or lower	70.4	29.6
Some college or associate's degree	94.6	5.4
Bachelor's degree or higher	95.5	4.5
Pearson <i>F</i> statistic	50.4*	
Age		
16–34	93.2	6.8
35–44	89.3	10.7
45–54	77.5	22.5
55–65	80.1	20.0
Pearson <i>F</i> statistic	12.1*	
Employment status		
Employed	88.0	12.0
Not employed	80.8	19.2
Pearson <i>F</i> statistic	7.7*	

\* $p < .05$ .

NOTE: Percentages represent the proportion of item respondents within the specified subgroup providing the specified response to the item. The Pearson *F* test determines whether a statistically significant relationship exists between the response to the item and education, household language, age, or employment status. Observations with missing data for a given variable are excluded. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

*This page is intentionally left blank.*

# Appendix C: Characteristics of Households

## Missing Auxiliary Data

As discussed in chapter 3, households that are missing data for a relatively large number of frame and commercial variables appear to show relatively low response rates, and to differ from households with more complete frame data with respect to some key estimates. This appendix provides a brief additional analysis of the characteristics of households that are missing a relatively large number of frame and commercial variables by linking the National Adult Training and Education Survey (NATES) sample to area-level estimates. Specifically, NATES-sampled households missing data for fewer than two key auxiliary variables are compared to those missing data for two or more with respect to tract-level 2008-2012 American Community Survey (ACS) estimates obtained from the 2014 Census Planning Database (PDB).<sup>30</sup>

As shown in table C.1, nearly all estimates show statistically significant differences between these two groups of households. On average, households missing data for two or more auxiliary variables are located in tracts with:

- Higher percentages of renters;
- Higher percentages of persons with less than a high school diploma;
- Higher percentages of persons below the poverty line;
- Higher percentages of persons ages 18 through 44;
- Higher percentages of non-White persons;
- Higher percentages of persons speaking a language other than English at home;
- Slightly lower employment rates;
- Higher percentages of persons who moved from another residence in the past year;
- Higher percentages of households in which no spousal relationship is present;
- Higher percentages of housing units in multi-unit structures; and
- Lower median household incomes.

These results therefore provide additional corroboration of prior findings, discussed by Harter et al. (2016), that neighborhood characteristics such as lower incomes and the presence of relatively mobile populations (including immigrants) tend to be associated with lower match rates to commercial databases.

---

<sup>30</sup> The comparison is conducted between households missing fewer than two variables and those missing two or more because this is the same cutoff used in the analysis discussed in section 3.2 of chapter 3. The 2008-2012 ACS estimates are used because these were the most recent estimates as of the date that the NATES data collection began.

**Table C.1. Selected tract-level estimates, by number of missing auxiliary variables: 2013**

Selected tract-level estimate	Mean		t statistic
	1 or fewer missing auxiliary variables	2 or more missing auxiliary variables	
Percent of housing units that are rented	30.1	38.4	14.7*
Percent of persons ages 25 or older without a high school degree	12.9	15.1	8.6*
Percent of persons below poverty level	13.1	16.7	12.1*
Percent of persons ages 18 through 44	35.0	37.2	7.9*
Percent non-White persons	30.9	36.4	7.5*
Percent of persons ages 5 or over speaking a non-English language at home	16.9	19.6	5.5*
Employment rate among civilians ages 16 or over	90.9	90.0	-6.4*
Percent of persons who moved from another residence in the past year	13.9	16.6	14.2*
Percent of households in which the householder moved in the year 2010 or later	9.9	11.8	14.5*
Percent of housing units constructed in 2010 or later	0.3	0.3	-1.0
Percent of housing units in which no spousal relationship is present	48.3	53.8	12.0*
Percent of persons who are female	51.0	50.8	-0.9
Percent of housing units in a multi-unit structure	27.2	36.5	16.5*
Median household income	61136.3	53755.6	-11.9*
Average household size	2.6	2.6	-5.9*

\*  $p < .05$

NOTE: Estimates are weighted using household-level selection weights.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Commerce, Bureau of the Census, American Community Survey (ACS), 2008–12.

Many of the variables shown in table C.1 are likely to be strongly correlated with each other. To account for this, table C.2 shows odds ratios from a logistic regression estimated on the NATES sample. The dependent variable for the regression is an indicator for households missing data for two or more key auxiliary variables. In the regression, only two tract-level estimates—the percent of persons who moved from another residence in the past year, and the percent of housing units that are in a multi-unit structure—show statistically significant coefficients. This suggests that, of the characteristics analyzed, indicators of population mobility and a high prevalence of multi-unit housing structures are the strongest independent predictors of missingness in auxiliary variables.



**Table C.2. Odds ratios for probability of having 2 or more missing auxiliary variables, by selected tract-level estimate: 2013**

Selected tract-level estimate	Odds ratio	Standard error	t statistic	95% confidence interval: lower bound	95% confidence interval: upper bound
Percent of housing units that are rented <sup>1</sup>	1.0	0.00	0.0	1.0	1.0
Percent of persons ages 25 or older without a high school degree <sup>1</sup>	1.0	0.01	1.1	1.0	1.0
Percent of persons below poverty level <sup>1</sup>	1.0	0.01	1.8	1.0	1.0
Percent of persons ages 18 through 44 <sup>2</sup>	1.0	0.01	-1.3	1.0	1.0
Percent non-White persons <sup>1</sup>	1.0	0.00	0.6	1.0	1.0
Percent of persons ages 5 or over speaking a non-English language at home <sup>2</sup>	1.0	0.00	-0.1	1.0	1.0
Employment rate among civilians ages 16 or over <sup>1</sup>	1.0	0.01	0.2	1.0	1.0
Percent of persons who moved from another residence in the past year <sup>1</sup>	1.0	0.01	2.4*	1.0	1.0
Percent of households in which the householder moved in the year 2010 or later <sup>1</sup>	1.0	0.01	1.3	1.0	1.0
Percent of housing units constructed in 2010 or later <sup>2</sup>	1.0	0.04	-0.5	0.9	1.1
Percent of housing units in which no spousal relationship is present <sup>1</sup>	1.0	0.00	0.4	1.0	1.0
Percent of persons who are female <sup>2</sup>	1.0	0.01	-1.1	1.0	1.0
Percent of housing units in a multi-unit structure <sup>1</sup>	1.0	0.00	4.6*	1.0	1.0
Median household income <sup>2</sup>	1.0	0.00	-0.4	1.0	1.0
Average household size	0.9	0.12	-0.7	0.7	1.2
Constant	1.4	1.36	0.4	0.2	9.6

\* p < .05

<sup>1</sup> Odds ratios are above 1.0 but round down to 1.0.

<sup>2</sup> Odds ratios are below 1.0 but round up to 1.0.

NOTE: Estimates are weighted using household-level selection weights.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Commerce, Bureau of the Census, American Community Survey (ACS), 2008–12.

## References

Harter, R., Battaglia, M.P., Buskirk, T.D., Dillman, D.A., English, N., Fahimi, M., Frankel, M.R., Kennel, T., McMichael, J.P., McPhee, C.B., Montaquila, J., Yancey, T., and Zukerberg, A.L. (2016). *Address-Based Sampling*. Oakbrook Terrace, IL: American Association for Public Opinion Research

*This page is intentionally left blank.*

## Appendix D: Standard Error Tables

**Table D.1. Standard errors for table 2.1: Percentage distribution of NATES respondents and eligible sample and estimated bias, by weighting type and selected auxiliary variables: 2013**

Selected auxiliary variable	With selection weights			With nonresponse-adjusted weights	
	Percent of eligible sample	Percent of respondents	Estimated bias	Percent of respondents	Estimated bias
Race/ethnicity stratum					
Black	1.08	1.15	0.32	1.18	0.34
Hispanic	0.69	0.70	0.21	0.73	0.24
Other	1.17	1.24	0.37	1.26	0.40
Ethnicity of the head of household					
Missing	0.78	0.87	0.50	0.86	0.50
White	0.98	1.04	0.52	1.06	0.55
Black	0.60	0.64	0.22	0.64	0.24
Hispanic	0.39	0.42	0.24	0.45	0.27
Asian/Pacific Islander	0.18	0.24	0.13	0.24	0.14
Other	0.22	0.27	0.15	0.25	0.15
Household income					
Missing	0.51	0.59	0.44	0.67	0.48
\$0–\$10,000	0.30	0.32	0.23	0.34	0.24
\$10,001–\$20,000	0.33	0.42	0.29	0.42	0.31
\$20,001–\$30,000	0.40	0.59	0.35	0.53	0.31
\$30,001–\$40,000	0.47	0.60	0.29	0.62	0.33
\$40,001–\$50,000	0.45	0.59	0.41	0.57	0.40
\$50,001–\$60,000	0.49	0.55	0.32	0.56	0.35
\$60,001–\$75,000	0.47	0.60	0.50	0.56	0.50
\$75,001–\$100,000	0.48	0.61	0.33	0.58	0.33
\$100,001–\$150,000	0.41	0.57	0.32	0.51	0.29
\$150,001+	0.25	0.33	0.19	0.30	0.18
Route type					
High rise	0.48	0.55	0.43	0.68	0.51
Street	0.65	0.82	0.48	0.83	0.50
Rural route	0.27	0.26	0.05	0.21	0.07
P.O. box	0.47	0.56	0.38	0.50	0.35
Education of the head of household					
Missing	0.81	0.89	0.58	0.91	0.60
Less than high school diploma	0.49	0.60	0.40	0.56	0.37
High school diploma	0.65	0.84	0.45	0.82	0.45
Some college	0.49	0.63	0.38	0.62	0.37
Bachelor's degree	0.41	0.49	0.30	0.45	0.29
Graduate degree	0.34	0.42	0.30	0.39	0.28

See notes at end of table.

**Table D.1. Standard errors for table 2.1: Percentage distribution of NATES respondents and eligible sample and estimated bias, by weighting type and selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	With selection weights			With nonresponse-adjusted weights	
	Percent of eligible sample	Percent of respondents	Estimated bias	Percent of respondents	Estimated bias
Age of the head of household					
Missing	0.92	0.96	0.52	1.04	0.58
18–24	0.13	0.13	0.12	0.15	0.14
25–34	0.33	0.47	0.25	0.45	0.25
35–44	0.59	0.76	0.42	0.72	0.38
45–54	0.52	0.64	0.36	0.63	0.35
55–65	0.63	0.73	0.37	0.69	0.37
Over 65	0.53	0.80	0.40	0.66	0.29
Census tract poverty rate					
20 percent or higher	0.80	0.94	0.42	0.96	0.46
Below 20 percent or missing	0.80	0.94	0.42	0.96	0.46
Phone number available in sampling frame					
Yes	0.89	1.02	0.53	0.97	0.53
No	0.89	1.02	0.53	0.97	0.53
NATES form type					
Individual	0.58	0.93	0.68	0.97	0.73
Booklet	0.58	0.93	0.68	0.97	0.73
Number of adults in household					
Missing	0.52	0.59	0.45	0.66	0.48
1	0.66	0.90	0.59	0.94	0.62
2	0.75	0.92	0.52	0.86	0.51
3	0.54	0.78	0.40	0.72	0.35
4	0.37	0.51	0.29	0.42	0.25
5	0.25	0.35	0.16	0.36	0.17
6	0.10	0.15	0.08	0.13	0.07
7	0.04	0.06	0.03	0.06	0.03
8	0.04	0.01	0.04	0.01	0.04
Home tenure					
Missing	0.61	0.69	0.45	0.77	0.50
Own	0.86	0.89	0.55	1.01	0.60
Rent	0.59	0.60	0.37	0.74	0.47
Census region					
Northeast	1.80	2.08	0.4	1.99	0.41
South	2.70	2.78	0.58	2.70	0.58
Midwest	2.23	2.44	0.50	2.38	0.50
West	2.08	2.24	0.58	2.23	0.59

See notes at end of table.

**Table D.1. Standard errors for table 2.1: Percentage distribution of NATES respondents and eligible sample and estimated bias, by weighting type and selected auxiliary variables: 2013—Continued**

Selected auxiliary variable	With selection weights			With nonresponse-adjusted weights	
	Percent of eligible sample	Percent of respondents	Estimated bias	Percent of respondents	Estimated bias
Dwelling type					
Missing	0.47	0.56	0.38	0.50	0.35
Single-family unit	0.54	0.75	0.49	0.78	0.52
Multi-unit	0.46	0.55	0.45	0.67	0.52
Gender of the head of household					
Missing	0.56	0.65	0.46	0.73	0.51
Female	0.56	0.77	0.51	0.79	0.52
Male	0.75	0.90	0.59	0.93	0.60
Marital status of the head of household					
Missing	0.69	0.72	0.54	0.76	0.57
Married	0.78	0.89	0.56	0.90	0.57
Single	0.55	0.65	0.41	0.71	0.45

NOTE: Percentages represent the proportion of eligible sampled households or respondent households with the specified characteristic. "Estimated bias" represents the difference between the respondent proportion and the eligible sample proportion.  
 SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

**Table D.2. Standard errors for table 2.2: Percentage distribution of NATES plus NRFU respondents and NATES-only respondents and estimated bias, by weighting type and selected reported characteristics: 2013**

Selected reported characteristic	With selection weights			With nonresponse-adjusted weights	
	Percent of NATES plus NRFU respondents	Percent of NATES-only respondents	Estimated bias	Percent of NATES-only respondents	Estimated bias
Has a professional certification or license					
Yes	0.85	0.91	0.91	0.86	0.88
No	0.85	0.91	0.91	0.86	0.88
Has a certificate					
Yes	0.77	0.77	0.80	0.74	0.80
No	0.77	0.77	0.80	0.74	0.80
Currently enrolled in college classes					
Yes	0.65	0.60	0.64	0.63	0.67
No	0.65	0.60	0.64	0.63	0.67
Highest educational attainment					
Less than high school diploma	0.52	0.52	0.44	0.51	0.44
High school diploma or equivalent	0.81	1.09	0.72	1.07	0.72
Some college or associate's degree	0.89	1.01	0.72	0.96	0.72
Bachelor's degree	0.69	0.79	0.53	0.78	0.53
Graduate or professional degree	0.64	0.69	0.52	0.64	0.51
Sex					
Male	0.86	0.86	0.78	0.84	0.82
Female	0.86	0.86	0.78	0.84	0.82
Age					
16–24	0.61	0.54	0.58	0.61	0.61
25–34	0.83	0.88	0.90	0.91	0.92
35–44	0.78	0.85	0.66	0.83	0.63
45–54	0.96	0.99	0.84	0.94	0.82
55–65	1.03	1.28	0.79	1.25	0.78
Race					
White	1.03	0.84	0.77	0.89	0.80
Black	0.90	0.71	0.68	0.78	0.72
Other or multiple	0.56	0.53	0.49	0.55	0.50
Origin					
Hispanic	0.97	0.64	0.78	0.70	0.78
Not Hispanic	0.97	0.64	0.78	0.70	0.78
Household income					
\$0–\$30,000	0.90	0.94	0.76	0.97	0.79
\$30,001–\$75,000	1.02	1.04	0.89	1.05	0.88
\$75,001+	0.87	1.02	0.86	1.00	0.85

NOTE: Percentages represent the proportion of NATES plus NRFU respondents or NATES-only respondents with the specified characteristic. "Estimated bias" represents the difference between the NATES-only proportion and the NATES plus NRFU proportion.  
 SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

**Table D.3. Standard errors for table 2.3: Percentage distribution of NATES respondents, by weighting type and selected reported characteristics: 2013**

<b>Selected reported characteristic</b>	<b>Selection-weighted percent</b>	<b>Nonresponse-adjusted percent</b>	<b>Change in estimate</b>
Completed an apprenticeship			
Yes	0.61	0.60	0.11
No	0.61	0.60	0.11
Completed other work-related training			
Yes	0.91	0.92	0.20
No	0.91	0.92	0.20
Speaks a language other than English at home			
Yes	0.82	0.89	0.17
No	0.82	0.89	0.17
English ability			
Speaks well or very well	1.43	1.33	0.37
Speaks not very well or not at all	1.43	1.33	0.37

NOTE: Percentages represent the proportion of NATES respondents with the specified characteristic. Observations with missing data for a given variable are excluded.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

*This page is intentionally left blank.*



# Appendix E: Sampling, Weighting, and Estimation Procedures

This appendix details the sampling, data collection, weighting, and estimation procedures for the two surveys on which this report is based: the National Adult Training and Education Survey (NATES), a mailed household survey sent to a sample of 10,000 addresses; and the NATES Nonresponse Follow-up Study (NRFU), an in-person survey of approximately 1,690 households that did not respond to the mailed NATES. Section E.1 reports sampling and data collection procedures. Section E.2 describes the creation of the household- and person-level weights used to generate the reported estimates. Section E.3 provides an overview of the variance estimation procedures, with an emphasis on the procedures for estimating the variance of the bias estimates reported in chapter 2. Section E.4 specifies the procedures and formulas used to calculate response rates to NATES and the NRFU. Finally, section E.5 discusses an important analytical consideration for comparisons of data from NATES and the NRFU.

## E.1 Sampling and Data Collection

Sampling and data collection for NATES and the NRFU were conducted by the U.S. Census Bureau. Table E.1 provides a calendar of data collection activities for both phases. The remainder of this section describes the sampling and data collection procedures for both the NATES and the NRFU.

**Table E.1. Timeline of NATES and NRFU data collection activities: 2013**

Activity	Date
Advance letters for NATES mailed	January 14, 2013
Initial set of questionnaires mailed	January 23, 2013
Reminder postcards mailed	February 4, 2013
Second set of questionnaires mailed	February 13, 2013
Third set of questionnaires mailed	March 6, 2013
Fourth set of questionnaires mailed	March 27, 2013
End of data collection for NATES	April 9, 2013
Advance letters for NRFU mailed	May 1, 2013
NRFU field collection activities begin	May 20, 2013
NRFU field collection activities end	July 1, 2013

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013; U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

## ***NATES household sampling***

The household-level target population for NATES was all U.S. households containing noninstitutionalized adults. The NATES sample was obtained from unused sample from the 2012 National Household Education Survey (NHES:2012). NHES:2012 had obtained a sample of 208,000 addresses from Marketing Systems Group (MSG), of which 160,000 were randomly selected for the 2012 data collection. The remaining 48,000 comprised the frame for the NATES sample. Several steps were followed to select the final NATES sample of 10,000 households.

First, since NHES:2012 had undersampled P.O. box addresses that were not the household's only way to get mail (OWGM), the file of 48,000 households had an oversample of these addresses. In order to reduce the proportion of non-OWGM P.O. boxes to correspond to the proportion in the original NHES:2012 frame, a random subsample of these addresses was selected for retention at a rate of 3 in 11. The remaining non-OWGM P.O. boxes were dropped. Approximately 37,750 addresses remained after this step.

Second, because the NATES questionnaire was not available in Spanish, the NHES:2012 oversample of the Hispanic stratum (Census tracts with 40 percent or more persons of Hispanic origin) was also removed. Addresses in the Hispanic stratum were subsampled for retention at a rate of 1 in 1.767335. The remaining households in the Hispanic stratum were deleted, leaving approximately 35,290 addresses.

Third, a random subsample of 1,200 households was selected for a separate pilot study intended to test different versions of the NHES household screener, leaving 34,090 available for NATES.

Fourth, the remaining households were restricted to include only addresses that were within the county-level Primary Sampling Units (PSUs) selected for the National Crime Victimization Survey (NCVS), a separate study sponsored by the U.S. Bureau of Justice Statistics. A PSU design was used to facilitate in-person data collection during the nonresponse follow-up phase of the study. Addresses not located in an NCVS PSU as of the summer of 2013 were dropped, leaving 21,690 addresses.

The NCVS PSU design included some PSUs selected with a probability of 1.0 (referred to as self-representing PSUs) and some selected with a probability of less than 1.0 (referred to as non-self-representing PSUs). In order to minimize travel costs in the nonresponse follow-up phase, the optimum design for NATES was to keep all 3,200 addresses in the non-self-representing PSUs and subsample the self-representing PSUs. Thus, 6,800 addresses in the self-representing PSUs were subsampled to obtain the final sample size of 10,000.

Households in the self-representing PSUs were subsampled as follows. The NHES sample already included an oversample of the Black stratum (Census tracts with 25 percent or more Black persons). The Other stratum (Census tracts not in the Black or Hispanic stratum) was further stratified by tract-level poverty rates. Specifically, the Other stratum was divided into a high-poverty stratum (tracts with poverty rates of 20 percent or higher) and a low-poverty stratum (tracts with poverty rates below 20 percent). The 820 households in self-representing PSUs and the high-poverty Other stratum were subsampled at a rate of 1 in 1.710692 while the remaining 17,670 households in self-representing PSUs were subsampled at a rate of 1 in 2.794401.

Together, the 3,200 households from the non-self-representing PSUs and the 6,800 households subsampled from the self-represented PSUs comprised the final NATES sample of 10,000 households.

Because all eligible members of sampled households were asked to fill out and return a NATES questionnaire, no within-household sampling was conducted.

### ***NATES data collection***

Data collection began with the mailing of advance notification letters to sampled addresses on January 14, 2013. The letters introduced the survey, informed the household that it had been selected to participate, and provided notice of the forthcoming questionnaire including the approximate time to complete the questionnaire. The letter also informed the household that it would receive a small token of appreciation. The letter included a toll-free number for the recipient to call with any questions. The advance letters and all NATES packages were addressed to “Dear Resident” in the mailing salutation. The packages were addressed to the “CURRENT RESIDENT.” All envelopes were preprinted with the Census Bureau logo on the left-hand side.

The initial topical packages were mailed to all sample addresses on January 23, 2013, and contained the following:

- A letter to the household that introduced the survey and requested that a questionnaire be filled out by each adult in the household;
- Either three single-person topical questionnaire booklets or one composite questionnaire that captured the data for three respondents in one booklet;
- A \$15 cash incentive in the form of three \$5 bills; and
- Either three pre-addressed, postage-paid return envelopes for those receiving three separate questionnaire booklets, or one pre-addressed, postage-paid return envelope for those receiving one composite questionnaire booklet.

Households with more than three adults were able to request additional questionnaires by calling the toll-free number on the cover letter. However, there were no separate mailings of additional questionnaires; rather, the questionnaires were sent with the next scheduled follow-up mailing.

A reminder postcard was sent to nonresponding households on February 4, 2013. Households that did not respond to the first mailing were sent topical packages in three subsequent mailings. Except in the third mailing, when most topical packages were shipped via FedEx, packages were shipped via U.S. Postal Service (USPS) First-Class mail.<sup>31</sup> Each follow-up mailing wave was sent 3 weeks after the previous follow-up wave to allow time for the receipt of completed questionnaires.

The follow-up packages included a cover letter, either three single-person topical questionnaire booklets or one composite questionnaire that captured the data for three respondents in one booklet, and a postage-paid return envelope. No incentive was included in any of the follow-up mailings.

The Census Bureau maintained a Telephone Questionnaire Assistance (TQA) hotline to assist respondents who called with questions about the questionnaire; address respondent concerns about confidentiality, purpose, sponsorship, and other similar issues; and convey the importance of survey participation to

---

<sup>31</sup> FedEx does not ship to P.O. boxes, so any packages in the third mailing with a P.O. box address were sent by USPS Priority Mail.

respondents who were reluctant to participate. Interviewers participating in NATES TQA were provided self-study training.

Respondents were encouraged to complete and mail back all forms sent to them in the pre-addressed, postage-paid return envelope addressed to the Census Bureau's main processing facility in Jeffersonville, Indiana. Upon receipt of the questionnaires, clerical staff immediately checked it into the Automatic Tracking and Control (ATAC) system and assigned a household-level outcome code. At this stage, a household received an outcome code of complete if it returned at least one questionnaire with at least one item answered. Additional outcome codes included refusals, blanks, duplicates, undeliverable as addressed (UAA), and various out-of-scope codes.

During data review, the Census Bureau conducted a second round of completeness checks for all returned questionnaires. At this stage, a questionnaire passed the completeness check if at least one of the following items was answered: highest education completed, sex, or age. Questionnaires that passed the completeness check were included in the final NATES data file, while questionnaires that failed the completeness check were excluded. However, for households for which all returned questionnaires failed the completeness check, the final household-level outcome code was not changed.

The NATES data collection ended on April 9, 2013. The hard cutoff date was necessary to ensure that the final sample for the NRFU follow-up interviews could be provided to the Census Bureau's Field Division for the allocation of interviewing workloads.

### ***NRFU household sampling***

The initial round of sampling for the NRFU took place on March 27, 2013, using the most current NATES outcome code at that time. In general, eligible households were those from which no form had been received, those from which a blank form had been received, and soft refusals. The Census Bureau planned to exclude hard refusals from the NRFU sample, but no NATES households were classified as hard refusals as of the sampling date. For the purpose of operational efficiency, several types of hard-to-reach addresses were classified as ineligible for the NRFU: those located in Alaska and Hawaii, those in PSUs with fewer than 3 expected NATES nonrespondents, those whose mailing addresses were not valid location addresses (e.g., P.O. boxes and rural routes), and those that presented great difficulty and expense in locating for a personal visit.

The sampling procedures differed between non-self-representing and self-representing PSUs. The non-self-representing PSUs—which were generally smaller and more geographically remote—were subsampled. The sampling probability for each non-self-representing PSU was inversely proportional to its original probability of selection for the NCVS. Specifically, PSUs whose original NCVS weight (WNCVS) was greater than 5.34041 were sampled with certainty, while those whose original NCVS weight was less than or equal to 5.34041 were sampled with probability  $\frac{WNCVS}{5.34041}$ . Within the non-self-representing PSUs that were selected, all NATES nonrespondents were included in the NRFU sample.

By contrast, all self-representing PSUs—which were generally larger with more concentrated populations—were retained for the bias study, but NATES nonrespondents within self-representing PSUs were subsampled with equal probability.

Out of the approximately 4,200 households that were eligible for the NRFU as of March 27, the initial round of sampling selected 1,800: approximately 620 from non-self-representing PSUs and approximately 1,180 from self-representing PSUs. However, sample units from which a NATES questionnaire was received after March 27 but prior to the final NATES cutoff on April 9 were counted as NATES respondents and thus dropped from the NRFU sample. Ultimately, approximately 410 households were dropped from the NRFU sample for this reason, leaving a sample of approximately 1,390. Because the anticipated response rate to the NRFU was approximately 60 percent, and the targeted number of completed interviews was 1,000, the decision was made to select an additional subsample of households that were NATES nonrespondents as of April 9. Thus, a second round of sampling on April 16 selected approximately 300 more households, all from self-representing PSUs. This led to a final NRFU sample size of 1,690.

### ***NRFU data collection***

The NRFU began with mail out of advance letters on May 1, 2013, and ended on July 1, 2013. A total of 214 Field Representatives participated in the NRFU. Field representatives and field supervisors who participated in the study completed four hours of self-study training.

The NRFU questionnaire consisted of a limited set of questions about household members and their educational experience. Specifically, the NRFU questionnaire was:

- A paper questionnaire;
- Designed to be answered by the person who answered the door at the sample address if he or she met the age requirement of 16 to 65;
- A matrix design to capture a short set of attributes about the people in the household; and
- Developed for the field representatives to read 12 questions to the respondents.

The NRFU questionnaire is shown in Appendix F.

Field representatives read the questionnaire to respondents, recorded the answers, and mailed the completed questionnaires to their Regional Office. The Regional Offices then sent completed questionnaires to the National Processing Center (NPC). All questionnaires were checked-in and processed at NPC using optical mark (OMR) and key from image (KFI) data capture technology.

A small number of households that were sampled for the NRFU returned NATES forms after April 9. These cases were assigned special outcome codes. Households that provided the interviewer with the questionnaire at the door, or that informed the interviewer that they had mailed a late form, were assigned a code of “205” (mailout questionnaire received). Households that were not interviewed in person, but from whom mailed questionnaires were received at the National Processing Center after April 9, were assigned a code of “250” (late mail return). In both of these scenarios, the data from the first NATES form received from the household was transposed to an NRFU form and included in the NRFU data file.<sup>32</sup>

---

<sup>32</sup> Note that this treatment of late mail returns applied only to households that were sampled for the NRFU. Data from late NATES forms returned after April 9 by non- NRFU households was not included in either the NATES or the NRFU data file, because, as of April 9, these households were nonrespondents that had not been selected for the NRFU.

The majority of households that responded to the NRFU required five or fewer contact attempts. Table E.2 shows the distribution of the final number of contact attempts among NRFU respondent households, as reported by the interviewer. Late mail returns are excluded from table E.2.

**Table E.2. Number of contact attempts required for NRFU respondent households: 2013**

<b>Number of contact attempts</b>	<b>Percent of NRFU respondent households</b>
0 or missing <sup>1</sup>	6.1
1 to 5	88.6
6 to 10	4.9
11 or more	0.5

<sup>1</sup> Refers to cases for which the interviewer indicated 0 contact attempts, or did not report the number of contact attempts, despite having completed the questionnaire.

NOTE: Number of contact attempts was reported by the interviewer. All proportions are unweighted. Late mail returns are excluded from proportions. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

Among the NRFU respondents who were interviewed in person, 40.5 percent stated that they had previously received the mailed NATES questionnaire, while 23.8 percent stated that they had not and 25.8 percent did not know. The remainder did not answer this item.

## **E.2 Weighting**

This section describes the procedures for calculating several sets of weights used in the analyses reported in this study. It first discusses the calculation of the NATES household-level selection weights (HBW) and the NATES household-level adjusted weights (HHW), which were applied to all NATES observations in the household-level analyses reported in chapters 2 and 3. It then discusses the calculation of the person-level selection weights (UPW) and the person-level adjusted weights (NPW) for NATES respondents, the latter of which were applied to all observations in the person-level analyses reported in chapter 4. Finally, it discusses the calculation of the final weights for NATES nonrespondent households that were sampled for the NRFU, which will be referred to as the NRFU weights (NRFUW); these weights were applied to all NRFU observations in the household-level analyses reported in chapters 2 and 3.

### ***NATES household-level selection weights***

The selection weight for each sampled household was the inverse probability of its selection for the final NATES sample. The calculation of each household's selection weight began with the household's original MSG weight (WMSG), the inverse probability of the household's being among the 208,000 households originally sampled by MSG for NHES:2012. WMSG was equal to 487.786827 for households in the Black stratum, 442.880128 for households in the Hispanic stratum, and 773.037833 for households in the Other stratum. Several adjustment factors were then applied to reflect the multiple rounds of subsampling described in section D.1:

- WSUB accounted for the probability of the household's being among the 48,000 households that were not included in the final sample for NHES:2012 and therefore available for sampling for NATES.
- WPOBNOWGM accounted for the removal of the oversample of P.O. boxes that were not the household's only way to receive mail.
- WHISP accounted for the removal of the oversample of the Hispanic stratum.
- WNONSCR accounted for the probability of the household's not being sampled for the separate NHES screener study.
- WNCVS represented the inverse probability of selection for each NCVS PSU used in the NATES study, thereby accounting for the restriction of the sample to households located within NCVS PSUs.
- WFINAL accounted for the subsampling of households within PSUs. Its value was determined by whether the household was located in a self-representing or non-self-representing PSU and, for households in the Other stratum, whether it was located in a high- or low-poverty tract.

$HBW_j$ , the household-level selection weight for household  $j$ , was the product of WT\_MSG and all of these adjustment factors:

$$HBW_j = WMSG_j * WSUB_j * WPOBNOWGM_j * WHISP_j * WNONSCR_j * WNCVS_j * WFINAL_j$$

Table E.3 provides the values of each adjustment factor for different types of households.

**Table E.3. Components of NATES household-level selection weights: 2014**

Component	Value
WMSG	If stratum = Black: 487.786827 If stratum = Hispanic: 442.880128 If stratum = Other: 773.037833
WSUB	If non-OWGM = Yes: 1.5 If non-OWGM = No: 5.51
WPOBNOWGM	If non-OWGM = Yes: 3.67 If non-OWGM = No: 1
WHISP	If stratum = Hispanic: 1.767335 If stratum = Black or Other: 1
WNONSCR	1.0351968
WNCVS	Varies by NCVS PSU
WFINAL	If SR = No: 1 If SR = Yes and stratum = Black or Hispanic: 2.794401 If SR = Yes and stratum = Other and poverty = High: 1.710692 If SR = Yes and stratum = Other and Poverty = Low: 2.794401

NOTE: non-OWGM = Yes refers to P.O. box addresses that are the only way for households to receive mail; non-OWGM = No refers to all other households. SR = Yes refers to households in self-representing NCVS PSUs; SR = No refers to households in non-self-representing NCVS PSUs. Poverty = High refers to households in Census tracts with poverty rates of 20 percent or higher; Poverty = Low refers to all other households.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

### ***NATES household-level adjusted weights***

The NATES selection weights were adjusted for household-level nonresponse using a standard procedure known as weighting class adjustment. On the basis of the household-level outcome codes as of April 9, 2013 (the cutoff date for the mailed NATES data collection), each sampled address was classified as a respondent (type R), a nonrespondent (type N), an ineligible case (type I), or a case of unknown eligibility (type U). Respondents were households with an outcome code of “01” (complete), with a small number of exceptions.<sup>33</sup> Nonrespondents were households with outcome codes of “03” (blank) or “05” (soft refusal); there were no hard refusals. Ineligible households were households with outcome codes of “10” or “20” through “36”, all of which correspond to various types of out-of-scope and undeliverable as addressed (UAA) statuses. Cases of unknown eligibility were households with an outcome code of “99”; these were cases for which no questionnaire was returned and no information on the eligibility of the address was obtained.

A procedure called Chi-Squared Automated Interaction Detection (CHAID) was then used to identify household-level characteristics associated with nonresponse. Because the characteristics used in this analysis needed to be available for both respondents and nonrespondents, the household-level CHAID model used a set of variables available in or linked to the NATES sampling frame. Table E.4 lists and defines the auxiliary variables used in the household-level CHAID model, along with an indication of whether each variable was determined by the procedure to be predictive of nonresponse. For variables for which values were missing for some households, “missing” was treated as its own category.

---

<sup>33</sup> Approximately 860 households had outcome codes of “01,” but were not included in the final NATES data file after the completeness check described in section D.1. Upon further examination, it was determined that approximately 850 of these households had indicated in the first two items on the NATES questionnaire that there were no eligible household members living at that address; these households were retained as complete cases because they had completed all parts of the questionnaire that were relevant to them. The remaining 10 households were reclassified as nonrespondents for the purpose of weighting and response rate calculation.



**Table E.4. Variables used in NATES household-level CHAID analysis: 2013**

Variable	Definition	Values	Predictive of nonresponse <sup>1</sup>	Missing rate <sup>2</sup>
Drop point	Whether the address is a single postal delivery point for multiple housing units	1 = Drop point 2 = Not a drop point 99 = Missing	No	0.0
Dwelling type	Whether the address is a single-family or multi-unit structure	1 = Single family 2 = Multi unit 99 = Missing	No	8.8
Phone match	Whether a phone number is available for the household on the sampling frame	1 = Matched 2 = Not matched	Yes	0.0
Address route type	Whether the address is a street address, P.O. box address, high-rise building address, or rural-route address	1 = Street 2 = High rise 3 = P.O. box 4 = Rural route	Yes	0.0
Seasonal address	Whether the address is seasonal	1 = Seasonal 2 = Not seasonal 3 = Educational seasonal	No	0.0
Address vacancy status	Whether the address is vacant	1 = Vacant 2 = Not vacant	No	0.0
Home tenure	Whether the address is owned or rented by the household	1 = Own 2 = Rent 99 = Missing	Yes	24.8
Only way to get mail	Whether a P.O. box address is the household's only address to get mail	1 = Only way to get mail 2 = Not only way to get mail	No	0.0
Educational attainment	Highest educational attainment of the head of household	1 = High school diploma 2 = Some college 3 = Bachelor's degree 4 = Graduate degree 5 = Less than high school diploma 99 = Missing	Yes	43.7

See notes at end of table.

**Table E.4. Variables used in NATES household-level CHAID analysis: 2013—Continued**

Variable	Definition	Values	Predictive of nonresponse <sup>1</sup>	Missing rate <sup>2</sup>
Race/ethnicity	Race or ethnicity of the head of household	1 = White 2 = Black 3 = Hispanic 4 = Asian/Pacific Islander 5 = Other 99 = Missing	Yes	40.1
Gender	Gender of the head of household	1 = Male 2 = Female 99 = Missing	Yes	22.0
Marital status	Marital status of the head of household	1 = Single 2 = Married 99 = Missing	No	34.0
Age	Age of the head of household	1 = 17 or younger 2 = 18-24 3 = 25-34 4 = 35-44 5 = 45-64 6 = 65 or higher 99 = Missing	Yes	42.4
Income	Household income	1 = \$10,000 or lower 2 = \$10,001-\$20,000 3 = \$20,001-\$30,000 4 = \$30,001-\$40,000 5 = \$40,001-\$50,000 6 = \$50,001-\$60,000 7 = \$60,001-\$75,000 8 = \$75,001-\$100,000 9 = \$100,001-\$150,000 10 = \$150,001 or higher 99 = Missing	Yes	19.3
Number of adults	Number of adults living in the household	1-8 = Number of adults in household 99 = Missing	Yes	19.5

<sup>1</sup> Indicates whether the variable was found by the household CHAID model to be predictive of household-level nonresponse.

<sup>2</sup> Indicates the unweighted percentage of NATES-sampled households for which information on the specified variable was not available on the sampling frame.

All sampled households were allocated to nonresponse adjustment cells defined by the characteristics identified by the CHAID analysis as being predictive of nonresponse. Table E.5 specifies the variables and values that defined the NATES household-level nonresponse adjustment cells.

**Table E.5. NATES household-level nonresponse adjustment cells: 2013**

Cell	Address route type	Income	Number of adults	Age	Gender	Phone match	Home tenure	Education	Ethnicity
1	3,4	1,99	†	†	†	†	†	†	†
2	3,4	2,10	2,3,4,6,7	†	†	†	†	†	†
3	3,4	2,10	1,5,99	†	†	†	†	†	†
4	2	†	†	6	†	†	†	†	†
5	2	†	†	4,5	1	†	†	†	†
6	2	†	†	4,5	2,99	†	†	†	†
7	2	3,6,7,8,9,10	†	2,3,99	†	†	†	†	†
8	2	1,5,99	†	2,3,99	†	†	†	†	†
9	2	2,4	†	2,3,99	†	†	†	†	†
10	1	†	†	†	†	1	2	†	†
11	1	5,7,9,10	2,5,6	6	†	1	1,99	†	†
12	1	5,7,9,10	1,3,4	6	†	1	1,99	†	†
13	1	1,2,3,4,6,8	†	6	†	1	1,99	1,2,3,4	†
14	1	1,2,3,4,6,8	†	6	†	1	1,99	5,99	†
15	1	1,9	†	5	†	1	1,99	†	†
16	1	6,10	†	5	†	1	1,99	†	†
17	1	2,4,7	†	5	†	1	1,99	†	†
18	1	3,5,8	2,5,7	5	†	1	1,99	†	†
19	1	3,5,8	1,3,4,6,8	5	†	1	1,99	†	†
20	1	†	2,4,6,7	3,4	†	1	1,99	†	†
21	1	†	1,3,5	3,4	†	1	1,99	†	†
22	1	†	†	2,7	†	1	1,99	†	2,99
23	1	†	†	2,7	†	1	1,99	†	1,3,4,5
24	1	†	†	4,5,6	†	2	2,99	†	†
25	1	3,6,9,10,99	†	2,3,99	†	2	2,99	†	†
26	1	1,2,4,5,7,8	†	2,3,99	†	2	2,99	†	†
27	1	†	†	3,6	†	2	1	†	†
28	1	†	†	5	†	2	1	3,4,99	†
29	1	†	†	5	†	2	1	2,5	†
30	1	†	†	5	†	2	1	1	†
31	1	7,10	†	2,4,99	†	2	1	†	†
32	1	5,8,9	2,3,4,6,7	2,4,99	†	2	1	†	†
33	1	5,8,9	1,5	2,4,99	†	2	1	†	†
34	1	1,2,3,4,6	†	2,4,99	†	2	1	†	†

† Not applicable (variable was not used to define the specified cell).

NOTE: Cells were defined using Chi-Squared Automated Interaction Detection (CHAID).

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

A household non-interview adjustment factor (HNIAF) was assigned to each adjustment cell using the following procedure. For cells with no type N nonrespondents, the HNIAF was set to 1.<sup>34</sup> For all other cells, the HNIAF for cell  $c$  was calculated as the inverse of the weighted  $ee$ -adjusted response rate within the cell:

$$HNIAF_c = \frac{\sum_{j \in R} HBW_{jc} + \sum_{j \in N} HBW_{jc} + ee_c * \sum_{j \in U} HBW_{jc}}{\sum_{j \in R} HBW_{jc}}$$

$$ee_c = \frac{\sum_{j \in R} HBW_{jc} + \sum_{j \in N} HBW_{jc}}{\sum_{j \in R} HBW_{jc} + \sum_{j \in N} HBW_{jc} + \sum_{j \in I} HBW_{jc}}$$

where:

$HBW_{jc}$  = the household-level selection weight;

$j$  = the household identifier;

$c$  = the nonresponse adjustment cell identifier;

$R$  = respondents;

$N$  = nonrespondents; and

$I$  = ineligible cases.

For all nonrespondent, ineligible, and unknown eligibility households, the household-level adjusted weight  $HHW$  was set equal to 0. For each respondent household  $j$ , the household-level adjusted weight was obtained by multiplying the household-level selection weight by the HNIAF for the household's adjustment cell  $c$ :

$$HHW_j = HBW_j * HNIAF_{jc}$$

### **NATES person-level selection weights**

As noted in section E.1, no within-household sampling was conducted for NATES; all eligible members of each sampled household were asked to fill out and return a questionnaire. However, because only 3 questionnaires were sent with each mailing wave, there were some households in which the number of eligible persons exceeded the number of available questionnaires. The person-level weighting procedure was designed to account for such situations by increasing the selection weights of person-level respondents in these households.

---

<sup>34</sup> In subsequent discussions with the U.S. Census Bureau, it was determined that the assignment of  $HNIAF = 1$  to all cells with no type N nonrespondents, even if the cell did contain cases of unknown eligibility (and thus had a response rate below 100 percent), was performed in error. Approximately 590 out of 5,470 NATES respondent households were in adjustment cells affected by this error. After this issue was discovered, the U.S. Census Bureau evaluated its likely impact on weighted percentage estimates, and determined that any impact on weighted survey estimates was likely to be minimal. Because of this, in light of the fact that NATES was a pilot study, NCES opted against further revisions to the data files. Readers are cautioned that NATES estimates in chapter 2 that use household-level nonresponse-adjusted weights, as well as those in chapter 4 that use person-level weights, may have been affected by this weighting error.

Therefore, in order to calculate person-level weights, it was necessary to first determine the number of persons in each household who were eligible for NATES. The first two items on the NATES questionnaire asked respondents to enumerate the eligible members of the household. Specifically, item QA asked for the number of household members ages 16 through 65. Respondents who entered 0 were directed to return the questionnaire without filling in any other items. Respondents who entered a number greater than 0 were directed to respond to QB, in which they were asked for the number of household members ages 16 through 65 who were no longer enrolled in high school. Respondents who entered 0 for QB were directed to return the questionnaire without filling in any subsequent items, while respondents who entered a number greater than 0 were asked to have each of those household members fill out and return the rest of the questionnaire.

In practice, however, there were numerous inconsistencies in the information reported in these two items. For example, some households returned a greater number of questionnaires than the numbers entered in QA and QB, and others reported a greater number in QB than QA even though it would be impossible for the number of individuals ages 16 through 65 and no longer in high school to exceed the total number of individuals ages 16 through 65. For this reason, a set of editing rules was used to determine  $m$ , the final number of eligible individuals in the household. Letting  $Q$  = the number of questionnaires received from the household,  $a$  = the number reported in QA, and  $b$  = the number reported in QB<sup>35</sup>:

- 1) If  $Q = a$ , then  $m = Q$
- 2) If  $Q \neq a$ , then  $m = \max(Q, b)$

The variables  $m$  and  $Q$  were then used to generate a person inflation factor (PIF) for each household. The PIF for each household was determined as follows:

- 1) If  $m \leq 3$ , then  $PIF = 1$
- 2) If  $m > 3$  and  $Q \leq 3$ , then  $PIF = m/3$
- 3) If  $m > 3$ ,  $Q > 3$ , and  $m \geq Q$ , then  $PIF = m/Q$ .<sup>36</sup>

For all person-level cases in the final NATES data file, each person's selection weight was calculated as the product of the household-level adjusted weight and the PIF for household  $j$ :

$$UPW_j = HHW_j * PIF_j$$

### ***NATES person-level adjusted weights***

In order to generate person-level nonresponse-adjusted weights (NPW), it was necessary to estimate the number of nonresponding persons within each household, based on the estimated number of eligible persons in the household as determined in the calculation of the person-level selection weights.

---

<sup>35</sup> For households that received individual questionnaire booklets rather than a composite booklet, every individual respondent was asked to fill out these first two items. For the purpose of this procedure, and for the calculation of within-household response rates (see section D.4), the responses to QA and QB were taken from the first person in each household to fill out a questionnaire. The procedure for selecting that case is described in section 5.2 of chapter 5.

<sup>36</sup> Note that, because  $m$  was defined as the maximum of  $Q$  and  $b$ , a situation in which  $m < Q$  would be impossible by construction.

This was done in several steps. First, the total number of eligible questionnaires received from the household was defined as  $E = Q - c$ , where  $c$  is the number of forms received from the household on which the individual reported that they were outside the eligible age range for NATES (16 through 65).<sup>37</sup> The proportion of questionnaires received from the household that were not outside the eligible age range was then calculated as  $p = E/Q$ . The following rules were then applied:

- 1) For households in which  $Q < 3$  and  $Q < m$ , it was assumed that there was within-household nonresponse, and the number of nonresponding persons in the household was estimated as  $N = p * (m - Q)$ .
- 2) For households in which  $Q \geq 3$  and/or  $Q = m$ , it was assumed that there was no within-household nonresponse, and  $N$  was set equal to 0.<sup>38</sup>

Once  $N$  was estimated for each household, a new record was created for each nonresponding person in order to create the input file for a second, person-level CHAID analysis. Because NATES did not include a screener stage, the only variables available for the person-level CHAID analysis were variables available in or linked to the NATES sampling frame. Table E.6 lists and defines the variables used in the person-level CHAID analysis, along with an indication of whether each variable was determined by the procedure to be predictive of person-level nonresponse.

**Table E.6. Variables used in NATES person-level CHAID analysis: 2013**

Variable	Definition	Values	Predictive of nonresponse <sup>1</sup>	Missing rate <sup>2</sup>
Stratum	Whether the household was located in the Black stratum, the Hispanic stratum, or the Other stratum	1 = Black 2 = Hispanic 3 = Other	Yes	0.0
Region	Whether the household was located in the Northeast, South, Midwest, or West Census region	1 = Northeast 2 = South 3 = Midwest 4 = West	Yes	0.0
Dwelling type	Whether the address is a single-family or multi-unit structure	1 = Single family 2 = Multi unit 99 = Missing	Yes	4.8
Home tenure	Whether the address is owned or rented by the household	1 = Own 2 = Rent 99 = Missing	Yes	15.3

See notes at end of table.

<sup>37</sup> Because the NATES questionnaire included no questions about whether an individual was currently enrolled in high school, it was impossible to determine individual eligibility on that criterion. This means that the number of eligible forms received is likely overestimated for some households.

<sup>38</sup> See note 36, above.

**Table E.6. Variables used in NATES person-level CHAID analysis: 2013—Continued**

Variable	Definition	Values	Predictive of nonresponse <sup>1</sup>	Missing rate <sup>2</sup>
Educational attainment	Highest educational attainment of the head of household	1 = High school diploma 2 = Some college 3 = Bachelor's degree 4 = Graduate degree 5 = Less than high school diploma 99 = Missing	Yes	34.0
Race/ethnicity	Race or ethnicity of the head of household	1 = White 2 = Black 3 = Hispanic 4 = Asian/Pacific Islander 5 = Other 99 = Missing	No	30.1
Age	Age of the head of household	1 = 17 or younger 2 = 18-24 3 = 25-34 4 = 35-44 5 = 45-64 6 = 65 or higher 99 = Missing	Yes	31.3
Income	Household income	1 = \$10,000 or lower 2 = \$10,001-\$20,000 3 = \$20,001-\$30,000 4 = \$30,001-\$40,000 5 = \$40,001-\$50,000 6 = \$50,001-\$60,000 7 = \$60,001-\$75,000 8 = \$75,001-\$100,000 9 = \$100,001-\$150,000 10 = \$150,001 or higher 99 = Missing	Yes	10.8
Number of adults	Number of adults living in the household	1-8 = Number of adults in household 99 = Missing	Yes	10.9

<sup>1</sup> Indicates whether the variable was found by the person CHAID model to be predictive of person-level nonresponse.

<sup>2</sup> Indicates the unweighted percentage of NATES person-level cases for which information on the specified variable was not available on the sampling frame.

All responding (R) and nonresponding (N) persons were allocated to nonresponse adjustment cells defined by the characteristics identified by the CHAID analysis as being predictive of nonresponse. Table E.7 specifies the variables and values that defined the NATES person-level nonresponse adjustment cells.

**Table E.7. NATES person-level nonresponse adjustment cells: 2013**

Cell	Age	Income	Number of adults	Stratum	Education	Region	Home tenure	Dwelling type
1	6	1,2,3,4,5,6,9	†	†	†	†	†	†
2	6	7,8,10	†	†	†	†	†	†
3	3	3,4,8,9,10	†	†	†	†	†	†
4	3	1,2,5,6,7	†	†	†	†	†	†
5	4	†	1,4,6	†	†	†	†	†
6	4	†	3,5	†	†	†	†	†
7	4	2,5,6,8	2	†	†	†	†	†
8	4	3,4,7,9,10	2	†	†	†	†	†
9	5	†	†	1	†	†	†	†
10	5	†	†	2	†	†	†	†
11	5	†	5,7	3	†	†	†	†
12	5	†	1,6	3	1,4,5	†	†	†
13	5	†	1,6	3	2,3,99	†	†	†
14	5	1,2,4,9	2	3	†	†	†	†
15	5	6,7	2	3	†	†	†	†
16	5	3,5,8,10	2	3	†	2,4	†	†
17	5	3,5,8,10	2	3	†	1,3	†	†
18	5	5,6	3,4	3	†	†	†	†
19	5	7,10	3,4	3	†	†	†	†
20	5	1,2,3,4,8,9	4	3	†	†	†	†
21	5	1,2,3,4,8,9	3	3	1,5,99	†	†	†
22	5	1,2,3,4,8,9	3	3	2,3,4	†	†	†
23	2,99	4	†	†	†	†	†	†
24	2,99	6,8	†	†	†	3,4	†	†
25	2,99	6,8	†	†	†	1,2	†	†
26	2,99	2,3,99	†	†	†	†	1	†
27	2,99	2,3,99	†	†	†	†	2,99	†
28	2,99	5,10	†	†	†	†	†	†
29	2,99	1,9,99	†	1	†	†	†	†
30	2,99	1,9,99	1	2,3	†	†	†	†
31	2,99	1,9,99	2,3,4,5,6,99	2,3	†	1	†	†
32	2,99	1,9,99	2,3,4,5,6,99	2,3	†	4	†	†
33	2,99	1,9,99	2,3,4,5,6,99	2,3	†	2,3	†	1
34	2,99	1,9,99	2,3,4,5,6,99	2,3	†	2,3	†	2,99

† Not applicable (variable was not used to define the specified cell).

NOTE: Adjustment cells were defined using Chi-Squared Automated Interaction Detection.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.



Each cell  $c$  was assigned a person non-interview adjustment factor (PNIAF) equal to the inverse of the weighted response rate within the cell:

$$PNIAF_c = \frac{\sum_{j \in R} UPW_{jc} + \sum_{j \in N} UPW_{jc}}{\sum_{j \in R} UPW_{jc}}$$

where:

$UPW_{jc}$  = the person-level selection weight;

$j$  = the person identifier;

$c$  = the nonresponse adjustment cell identifier;

$R$  = respondents; and

$N$  = nonrespondents.

For all nonresponding persons, as well as all responding persons who reported that their age was outside the eligible age range, the person-level adjusted weight NPW was set equal to 0. For each eligible responding person  $j$ , the person-level adjusted weight was obtained by multiplying the person-level selection weight by the PNIAF for the person's adjustment cell  $c$ :

$$NPW_j = UPW_j * PNIAF_{jc}$$

It is important to note that, relative to a person-level weighting procedure for a two-stage survey with a screener roster and within-household sampling, the person-level weighting procedure for NATES required a number of assumptions in order to determine the number of eligible persons and the number of nonrespondents in each household. This is attributable to two characteristics of the NATES data collection. First, rather than being asked to provide a roster of household members, households were asked to simply provide the number of eligible household members. As noted above, inconsistencies and errors in the responses to these questions were found for a number of households. Second, rather than using a random procedure to sample individual eligible persons from within each household to receive the topical questionnaire, the single-stage design used for NATES required households to determine for themselves who was eligible to take the survey and ensure that all such individuals returned a questionnaire. Both of these factors could have led to errors in the estimation of the number of person-level eligible cases and nonrespondents in the NATES sample.

### ***NRFU weights***

To generate final weights for the NATES nonrespondent households that were sampled for the NRFU, the original household selection weights needed to be multiplied by an NRFU inflation factor (NIF) equal to the inverse probability of selection for the NRFU. Because the sampling procedures differed between non-self-representing and self-representing PSUs, the calculation of the NIF also differed.

Within non-self-representing PSUs, there was subsampling of PSUs but no within-PSU subsampling. The weighting procedure for households in these PSUs began with WNCVS, defined above as the original NCVS weight for each PSU. As noted in section D.1, non-self-representing PSUs for which WNCVS > 5.34041 were sampled with certainty for the NRFU, and all households in these PSUs were included in

the final NRFU sample. Therefore, the NIF for households in these PSUs was set equal to 1. Non-self-representing PSUs for which  $WNCVS \leq 5.34041$  were sampled with probability  $\frac{WNCVS}{5.34041}$ , and all households in the sampled PSUs were included in the final NRFU sample. Therefore, the NIF for households in these PSUs was set equal to  $\frac{5.34041}{WNCVS}$ .

Within self-representing PSUs, there was no subsampling of PSUs, but households were sampled within PSUs. As noted in section D.1, there were two rounds of sampling within the self-representing PSUs: one on March 27 and one on April 16. The NIF for households selected in the first round of sampling was set equal to 2.56768, which was the inverse probability of selection from among the households in self-representing PSUs that were eligible for the NRFU as of March 27. The NIF for households selected in the second round of sampling was set equal to 8.013517; this was the inverse probability of selection from among households that were eligible for the NRFU as of April 9 and had not been selected in the first round, multiplied by the inverse probability of rejection from the first round.

$NRFUW_j$ , the final NRFU weight for household  $j$ , was the product of the household's original household-level selection weight and its NIF:

$$NRFUW_j = HBW_j * NIF_j$$

The NRFU weights were not adjusted for nonresponse to the NRFU interviews. In addition, because only one person in each household was interviewed for the NRFU, no further person-level adjustments were necessary for the NRFU weights.

## E.3 Variance Estimation

### **Overview of variance estimation procedures**

In surveys with complex sample designs, direct estimates of standard errors typically underestimate the variability in the estimates (Wolter 1985). The NATES and the NRFU sample design and weighting included procedures that deviated from the assumption of simple random sampling, such as oversampling in areas with higher concentrations of Blacks and in high-poverty areas, and sampling households within PSUs with differential sampling probabilities. In order to reflect these aspects of the sample design and weighting, the standard errors of all estimates reported in this study were calculated using a jackknife replication procedure.

Replication involves splitting the entire sample into a set of groups, or replicates, based on the actual sample design of the survey. The survey estimates can then be computed for each replicate by creating replicate weights that mimic the actual sample design and estimation procedures used in the full sample. The variation in the estimates computed from the replicate weights can then be used to directly estimate the sampling errors of the estimates from the full sample.

To create replicate weights, the initial NATES sample of 10,000 households was divided into 70 random subsamples. The addresses were then assigned 70 replicate selection weight variables (REPBW1 through REPBW70) on the basis of the following procedures. REPBW1 was created by multiplying the full-sample selection weight (HBW) by 0 if the household was in the first subsample and  $\frac{70}{69}$  otherwise.

Similarly, REPBW2 was created by multiplying HBW by 0 if the household was in the second subsample and  $\frac{70}{69}$  otherwise. This procedure was repeated to create all 70 household-level replicate selection weights.

The household-level replicate selection weights were then multiplied by the adjustment factors defined in section D.2 to generate household-level replicate adjusted weights (HHW1-HHW70), person-level replicate selection weights (UPW1-UPW70), person-level replicate adjusted weights (NPW1-NPW70), and household-level replicate NRFU weights (NRFUW1-NRFUW70). Specifically, for each household  $j$ :

$$HHWn_j = REPBWn_j * HNIAF_{jc}$$

$$HHWn_j = REPBWn_j * HNIAF_{jc}$$

$$NPWn_j = UPWn_j * PNIAF_{jc}$$

$$NRFSWn_j = REPBWn_j * NIF_j$$

where  $n = 1, 2, \dots, 70$

Using the 70 jackknife replicate weights, the variance  $v(\hat{\theta})$  of an estimate  $\hat{\theta}$  can be estimated as:

$$v(\hat{\theta}) = \frac{69}{70} \sum_{i=1}^{70} (\hat{\theta}_i - \hat{\theta})^2$$

where:

$\hat{\theta}_i$  = the estimate calculated using the  $i^{\text{th}}$  replicate weight

$\hat{\theta}$  = the estimate calculated using the full-sample weight

The standard error  $se(\hat{\theta})$  is the square root of the variance.

#### *Variance of bias*

A variant of the above formula was used to calculate the standard errors of the bias estimates presented in chapter 2. The variance of the difference between two estimated proportions  $\hat{p}$  and  $\hat{q}$  is given by:

$$v(\hat{p} - \hat{q}) = \frac{69}{70} \sum_{i=1}^{70} [(\hat{p}_i - \hat{q}_i) - (\hat{p} - \hat{q})]^2$$

The standard error of the difference is the square root of the variance of the difference. This formula was used to calculate the standard error of the bias estimates in tables 2.1 through 2.3 by substituting the appropriate values for  $\hat{p}$  and  $\hat{q}$ . Specifically, adopting the terminology used in chapter 2:

- In table 2.1,  $\hat{p}$  represents the selection-weighted or nonresponse-adjusted respondent proportion, while  $\hat{q}$  represents the selection-weighted eligible sample proportion.
- In table 2.2,  $\hat{p}$  represents the selection-weighted or nonresponse-adjusted NATES-only proportion, while  $\hat{q}$  represents the selection-weighted NATES plus NRFU proportion.
- In table 2.3,  $\hat{p}$  represents the nonresponse-adjusted NATES-only proportion, while  $\hat{q}$  represents the selection-weighted NATES-only proportion.

## E.4 Response Rate Calculation

The final response rate to the mailed NATES was the product of the household-level response rate and the person-level response rate. This section describes the procedures used to calculate first the household-level response rate and then the person-level response rate to NATES. It then describes the separate procedures used to calculate the final response rate to the in-person NRFU.

### ***NATES household-level response rate***

As was the case with the development of the NATES household-level adjusted weights (see section E.2), the response status of each sampled household was determined by its outcome code as of April 9, 2013 (the cutoff date for the mailed NATES data collection). Respondents (R) were households with an outcome code of “01” (complete), with a small number of exceptions.<sup>39</sup> Nonrespondents (N) were households with outcome codes of “03” (blank) or “05” (soft refusal). Ineligible households (I) were households with outcome codes of “10” or “20” through “36”, all of which correspond to various types of undeliverable as addressed (UAA) statuses. Cases of unknown eligibility (U) were households with an outcome code of “99”; these were cases for which no questionnaire was returned and no information on the eligibility of the address was obtained.

The household-level response rate for the mailed NATES was calculated using the American Association for Public Opinion Research (AAPOR) response rate 3 (RR3) formula (AAPOR 2015):

$$HRR = \frac{\sum_{j \in R} HBW_j}{\sum_{j \in R} HBW_j + \sum_{j \in N} HBW_j + ee * \sum_{j \in U} HBW_j}$$

$$ee = \frac{\sum_{j \in R} HBW_j + \sum_{j \in N} HBW_j}{\sum_{j \in R} HBW_j + \sum_{j \in N} HBW_j + \sum_{j \in I} HBW_j}$$

where  $HBW_j$  = the household-level selection weight for household  $j$

---

<sup>39</sup> See note 34, above.

Table E.8 provides the approximate unweighted count of sampled addresses with each household-level outcome code as of the April 9, 2013 cutoff.

**Table E.8. NATES household-level outcome codes, definition of each outcome code, and number of sampled households with each outcome code: 2013**

Outcome code	Definition	Number of sampled households
01	Complete <sup>1</sup>	5,480
03	Blank	60
05	Soft refusal	10
10	Out of scope	30
20	UAA with address correction	10
21	Not deliverable as addressed	150
22	Insufficient address	30
23	Moved, left no address	10
24	Unclaimed	#
25	Attempted - not known	40
26	No such street	#
27	No such street number	30
28	Vacant	560
30	No mail receptacle	20
31	P.O. box closed - no forwarding order	20
33	Deceased	10
34	Forwarding order has expired	#
36	UAA missing unit/apartment designation	10
99	Mailed, not yet returned	3,550
<b>Total</b>		<b>10,000</b>

# Rounds to zero.

<sup>1</sup> Approximately 860 households had outcome codes of "01," but were not included in the final NATES data file after the completeness check described in section D.1. Upon further examination, it was determined that approximately 850 of these households had indicated in the first two items on the NATES questionnaire that there were no eligible household members living at that address; these households were retained as complete cases because they had completed all parts of the questionnaire that were relevant to them. The remaining 10 households were reclassified as nonrespondents for the purpose of weighting and response rate calculation.

NOTE: Figures represent the unweighted count of sampled households with the specified outcome code as of April 9, 2013, the cutoff date for the NATES data collection. Counts are rounded to prevent disclosure of restricted-use information. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

### ***NATES person-level and final response rates***

For each household that returned at least one questionnaire that passed the completeness check described in section E.1, the number of eligible persons and the number of nonrespondents in each household were estimated using the same procedure as in the development of person-level base and adjusted weights (see

section E.2). Because there were no person-level cases of unknown eligibility, the use of an *ee* adjustment factor was unnecessary. Therefore, the person-level response rate to the mailed NATES questionnaire was calculated as the sum of the person-level selection weights for responding persons, divided by the sum of the person-level selection weights for responding and nonresponding persons:

$$PRR = \frac{\sum_{j \in R} UPW_j}{\sum_{j \in R} UPW_j + \sum_{j \in N} UPW_j}$$

where  $UPW_j$  = the person-level selection weight for person  $j$

The final response rate to the mailed NATES questionnaire was calculated as the product of the household- and person-level response rates:

$$RR = HRR * PRR$$

Table E.9 provides the approximate unweighted count of respondents, nonrespondents, and ineligible cases at the person level.

**Table E.9. NATES person-level outcomes: 2013**

<b>Outcome</b>	<b>Number of persons</b>
Respondent	7,540
Nonrespondent <sup>1</sup>	290
Ineligible	530
<b>Total</b>	<b>8,360</b>

<sup>1</sup> The number of person-level nonrespondents was estimated using questionnaire information on the number of eligible persons in each household.

NOTE: Counts are unweighted and are rounded to prevent the disclosure of restricted-use information. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey (NATES), 2013.

### ***NRFU response rate***

The NRFU response rate represents the weighted proportion of households sampled for the NRFU from whom usable data were collected between April 9 and July 1. It is important to emphasize that all households sampled for the NRFU were nonrespondents to the mailed NATES as of the April 9 cutoff; therefore, in the remainder of this section, references to respondents, nonrespondents, and ineligible cases represent the outcomes of the in-person NRFU data collection, not the mailed NATES data collection.

Households sampled for the NRFU were classified as respondents (R), nonrespondents (N), and ineligible cases (I) on the basis of their final NRFU outcome code. There were no cases of unknown eligibility. In general, respondents were households with an outcome code of “201” (interview), “205” (mailout questionnaire received), or “250” (late mail return). Nonrespondents were households with an outcome code of “213” (language barrier), “217” (temporarily absent), or “218” (refusal). Ineligible cases were households with an outcome code of “226” (vacant), “247” (multi-unit address without unit designation),

“248” (out-of-scope), and “249” (unable to locate/bad address). The only exception to these rules was that, if a household had a respondent outcome code but the interviewee reported being outside the eligible age range (16 through 65), the household was reclassified as ineligible.

The final response rate to the in-person NRFU was calculated as the sum of the NRFU weights for respondents, divided by the sum of the NRFU weights for respondents and nonrespondents:

$$NRR = \frac{\sum_{j \in R} NRBSW_j}{\sum_{j \in R} NRBSW_j + \sum_{j \in N} NRBSW_j}$$

where  $NRFUW_j$  = the NRFU weight for household  $j$

Table E.10 provides the approximate unweighted count of NRFU-sampled households with each outcome code.

**Table E.10. NRFU outcome codes, definition of each outcome code, and number of sampled households with each outcome code: 2013**

Outcome code	Definition	Number of sampled households
201	Interview <sup>1</sup>	1,130
205	Mailout questionnaire received <sup>1</sup>	10
213	Language barrier	10
217	Temporarily absent	50
218	Refusal	180
226	Vacant	130
247	Multi unit address without unit designation	10
248	Out of scope	80
249	Unable to locate/bad address	20
250	Late mail return <sup>1</sup>	70
<b>Total</b>		<b>1,690</b>

<sup>1</sup> Approximately 30 households that had a respondent outcome code (201, 205, or 250), but for which the interviewed respondent reported being younger than 16 or older than 65, were reclassified as out of scope for the purpose of analysis and response rate calculation.

NOTE: Figures represent the unweighted count of NRFU-sampled households with the specified outcome code as of the end of the NRFU field period. Counts are rounded to prevent disclosure of restricted-use information. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU), 2013.

### **Combined NATES and NRFU response rate**

The combined response rate to NATES and the NRFU was calculated using the AAPOR (2015) standard for surveys that subsample nonrespondents. Because the bias analysis was conducted at the household

level, this response rate was calculated at the household level. In this procedure, the “combined” response status was defined as follows:

- Respondents (R) consisted of NATES respondents, *and* NATES nonrespondents and unknown eligibles that were sampled for and responded to the NRFU.
- Nonrespondents (N) consisted of NATES nonrespondents and unknown eligibles that were sampled for the NRFU but then classified as nonrespondents to the NRFU.
- Ineligibles (I) consisted of NATES nonrespondents and unknown eligibles that were sampled for the NRFU but then classified as ineligible at the NRFU stage.
- NATES nonrespondents and unknown eligibles that were not sampled for the NRFU were dropped from the calculation.

The weight used in the final response rate calculation was a combined weight (COMBWGT) equal to the NATES selection weight (HBW) for NATES cases and the NRFU selection weight (NRFUW) for the NRFU cases. This accounted for the subsampling of cases for the NRFU. Because there were no unknown eligibles as of the completion of the NRFU data collection, the combined response rate was calculated as of the sum of the combined weights for respondents, divided by the sum of the combined weights for respondents and nonrespondents:

$$CRR = \frac{\sum_{j \in R} COMBWGT_j}{\sum_{j \in R} COMBWGT_j + \sum_{j \in N} COMBWGT_j}$$

where  $COMBWGT_j$  = the combined weight for household  $j$

## E.5 Combining NATES and NRFU Data

Aside from the mode difference and the use of a shortened questionnaire, a key design difference between the mailed NATES and the NRFU was that the in-person interviewers did not attempt to collect data from all eligible household members. Rather, the interview was administered to the person who answered the door, if that person was within the target population; or, if not, to the most readily available eligible person. Thus, the NRFU used a non-random convenience method to select a single eligible member of each household.

This inherent difference in the design of the two stages raised a complication for comparisons between NATES respondents and NRFU respondents. Because the mailed NATES attempted to collect data from all eligible members of each household, data from NATES respondents could, in principle, be used to create person-level estimates.<sup>40</sup> However, for the in-person NRFU, data were available only from a single “most convenient” respondent within each household. In surveys in which a single person is randomly selected from each household (e.g., full-scale NHES administrations), a standard method of creating person-level estimates is to multiply the household-level selection weight by the number of eligible

---

<sup>40</sup> The term “person-level estimates” is used here to refer to population estimates for which the population of eligible persons is the denominator—for example, the percent of persons who hold a particular educational credential. The term “household-level estimates” is used to refer to population estimates for which the population of eligible households is the denominator—for example, the percent of households in which the “most convenient” respondent holds a particular educational credential.



persons in the household, obtaining a person-level selection weight. However, this method implicitly assumes that a random or at least quasi-random method was used to select the respondent, with every eligible household member having an equal probability of selection. Because the NRFU respondent was not selected randomly, and the characteristics of the “most convenient” respondent are likely to differ systematically from other household members (cf. Lavrakas 2008), the application of a weighting adjustment for within-household selection would not have been a valid method of obtaining person-level estimates for the NRFU. For this reason, the NRFU data (when weighted by household-level selection weights) are properly interpreted as providing household-level estimates. For example, the proportion of NRFU respondents reporting a certification or license could not be interpreted as the estimated proportion of persons holding a certification or license, but rather as the estimated proportion of households in which the “most convenient” responder holds a certification or license.

If left uncorrected, this difference between the NATES and NRFU designs could have confounded interpretations of differences between the NATES and NRFU proportions. If the NRFU proportions (with or without a weighting adjustment for within-household selection) were compared to person-level NATES proportions—that is, proportions estimated using all household members who returned a mailed questionnaire—some of the difference between the proportions could be attributable not to nonresponse bias but rather to the difference in the population of which the estimates are representative. In essence, estimates of nonresponse bias would be confounded by the fact that the NRFU proportions are likely to be subject to within-household selection bias while the person-level NATES proportions are not or at least are assumed to not be.

For this reason, an attempt was made to improve the comparability of the NATES and NRFU proportions for the analysis of nonresponse bias by identifying the equivalent to the “first, most convenient” responder for households that returned more than one mailed NATES questionnaire. This was done by retaining the data from the questionnaire that appeared to have been completed first. For households that received the single booklet, this was the individual who filled out the questionnaire closest to the front of the booklet. For households that received three booklets, this was the individual who filled out the lowest-numbered questionnaire. For households that returned mailings on more than one date, data were retained from the earliest mailing. The resulting household-level NATES dataset was used for all analyses that combined or compared NATES and NRFU data—specifically, the bias analysis in chapter 2 and the regressions in chapter 3. All such comparisons were weighted using household-level selection weights to account for complex aspects of the NATES sampling and NRFU subsampling design.

In essence, this analytic approach aims to reduce a known confounding factor in comparisons between the NATES and NRFU proportions by introducing a semi-comparable selection bias into the NATES proportions. For this reason, the estimates presented in this report may not be comparable to estimates of similar characteristics (e.g., credential prevalence) obtained from other nationally representative surveys. It is also important to note that, because no household roster was collected, this approach implicitly relies on the assumption that the person who filled out the “first” mailed questionnaire would have been the person to respond to the in-person interview had the household been sampled for the NRFU, which cannot be empirically verified. To the extent that this assumption is inaccurate, estimates of nonresponse bias may still be confounded by differences in selection bias.

## References

- Lavrakas, P. J. (2008). *Within-household respondent selection: An MRC staff point of view*. New York, NY: Media Rating Council.
- U.S. Department of Education, National Center for Education Statistics. (2012). *National Household Education Survey (NHES)*. Data available at <http://nces.ed.gov/nhes/dataproducts.asp>.
- U.S. Department of Education, National Center for Education Statistics. (2013). *National Adult Training and Education Survey (NATES)*. [no publicly available dataset]
- U.S. Department of Education, National Center for Education Statistics. (2013). *National Adult Training and Education Survey Nonresponse Follow-up Study (NATES NRFU)*. [no publicly available dataset]
- U.S. Department of Justice, Bureau of Justice Statistics. (2013). *National Crime Victimization Survey (NCVS)*. Data available at <http://www.bjs.gov/developer/ncvs/index.cfm>.
- Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

## **Appendix F: National Adult Training And Education Survey (NATES) Questionnaire**

## The 2013 National Adult Training and Education Survey



251 03011

Conducted by

**UNITED STATES DEPARTMENT OF COMMERCE**  
Economics and Statistics Administration  
U.S. Census Bureau



**NATES-10AC**  
(01/03/2013)



---

---

## Instructions

- ◆ The Department of Education is studying the education and job training experiences of adults and youth. Each household is different, and we need your response.
  - ◆ Each person living at this address, ages 16 to 65 should fill out a survey. Youths who are still in high school should not fill out a survey.
  - ◆ Surveys should be completed by adults and youth who are temporarily away from home (for example, on vacation) but do not include adults or youth who are living at another address for an extended period of time (for example, living in college dormitories).
  - ◆ If you need additional surveys, please call us on our toll-free number: 1-800-845-8243.
  - ◆ Return each completed survey using the postage-paid envelopes provided.
  - ◆ To answer a question, simply mark the box [X] that best represents the answer.
  - ◆ Please use a black or blue pen, if available.
- 
- 



## Introduction

### Start Here

Adults acquire their job skills in many ways, including formal education, on-the-job-training, and other work training. This survey asks about all of these, including sections on professional certifications and licenses, educational certificates and other education degrees and classes, apprenticeship programs, and other classes taken for work reasons.

You will be asked to answer only the sections that apply to you. Please start with question A below. (Only one adult in the household needs to answer questions A and B on this page.)

**A. How many of the people living in this household are ages 16-65?**

If no one in this household is ages 16-65, please enter "0" in the box and return the questionnaire in the postage-paid envelope. It is important that you return your questionnaire. No one in your household needs to complete any other questionnaires.

**B. Of these people ages 16-65, how many are no longer in high school?**

☐

If everyone in this household ages 16-65 is still in high school, please mark this box and return this questionnaire in the postage-paid envelope. No one in your household needs to complete any other questionnaires.



*Please have each of these people fill out a questionnaire.*

***If you have any questions or need additional questionnaires, please contact us toll-free at 1-800-845-8243.***

251 03037



## Educational Attainment

1. What is the highest degree or level of school you have completed? (Mark one.)

☐ Elementary or high school, but no high school diploma or GED

GO TO question 4

☐ High school diploma, GED, or other high school completion

☐ Some college credit but less than one year of college credit

GO TO question 3

☐ 1 or more years of college credit, no degree

☐ Associate's degree (for example, AA, AS)

☐ Bachelor's degree (for example, BA, BS)

☐ Master's degree (for example, MA, MS, MEd, MEd, MSW, MBA)

☐ Professional degree beyond a bachelor's degree (for example, MD, DDS, DVM, LLB, JD)

☐ Doctorate degree (for example, PhD, EdD)

2. What was the major or field of study for your highest level of education? If there was more than one, please choose the one you consider most important.

Write in:

3. Did you complete your high school requirements through a regular high school diploma, or through the GED or other high school equivalency? (Mark one.)

☐ Regular high school diploma

☐ GED or other high school equivalency

4. Do you have a professional certification or a state or industry license? A professional certification or license shows you are qualified to perform a specific job and includes things like Licensed Realtor, Certified Medical Assistant, Certified Construction Manager, a Project Management Professional certification, or an IT certification.

☐ Yes

☐ No

GO TO question 20

5. Thinking of all the certifications and licenses you have, did you get any of them for work-related reasons, or were they all for personal interest? (Mark one.)

☐ I got ONE OR MORE certifications or licenses for work-related reasons

GO TO question 6

☐ I did NOT GET ANY certifications or licenses for work-related reasons

GO TO question 20

► Continue on the next page.

251 03045

NATES-10AC

## Certification and Licensure

➤ In the questions below, we ask a few details about the certification or license that you most recently earned for work-related reasons.

➤ Please answer these questions only about this MOST RECENT work-related certification or license. If you got a certification as part of getting a license, please respond for the license.

**6. In what year were you first issued your MOST RECENT work-related certification or license?**

Write in year:

**7. Who issued this certification or license? (Mark one.)**

- ☐ Federal, state, or local government
- ☐ Professional or trade association (for example, Pediatric Nursing Certification Board, National Exercise and Sports Trainers Association, CompTIA)
- ☐ Business or company (for example, Microsoft™, 3M Company™, Xerox®)
- ☐ Other group or organization (specify) ↴

**8. Why did you get this certification or license? (Mark "Yes" or "No" for each.)**

	Yes ▼	No ▼
To get a job in a new field . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To get a promotion or raise in pay . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To stay current in my field or expand skills in my field . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To start my own business . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To meet an employer requirement . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
Other (specify) ↴ . . . . .	<input type="checkbox"/>	<input type="checkbox"/>

**9. Did you have to pass a test or exam or demonstrate your skills to get this certification or license?**

- ☐ Yes
- ☐ No

**10. What kind of courses, training, or instruction (online or in-person) did you take in order to prepare for this certification or license? (Mark all that apply.)**

- ☐ I did not need any courses, training, or instruction
- ☐ I took vocational or occupationally focused high school courses
- ☐ I took courses from a vocational or trade school, community or technical college, or other college or university
- ☐ I took courses from a private company or my employer
- ☐ I participated in on-the-job training, an internship, or an apprenticeship
- ☐ I studied on my own
- ☐ Other (specify) ↴

**11. Do you have to earn continuing education units (CEUs) or other professional development credits to maintain this certification or license?**

- ☐ Yes
- ☐ No ➔

**GO TO question 14**

► **Continue on the next page.**

251 03052





12. Which ONE of the following best describes the **MOST RECENT** activity you engaged in to earn your continuing education or other professional development credits for this certification or license? (Mark one.)

☐ Have not yet had to meet these requirements



**GO TO question 14**

☐ Attended conference or demonstration (online or in-person)

☐ Completed class or seminar (online or in-person)

☐ Read instructional materials (online or hardcopy)

☐ Other (specify)

13. Who was the main provider of the instruction or learning materials for the activity you indicated in question 12? (Mark one.)

☐ My employer

A group other than my employer:

☐ Professional or trade association

☐ Labor union or labor organization

☐ Community or technical college, vocational or trade school, college, or university

☐ Federal, state, or local government

☐ Private training company

☐ Other (specify)

14. What is the name of your **MOST RECENT** certification or license? Please do not use abbreviations.

Write in:

15. What kind of work is this certification or license for? (for example: teaching, vocational nursing, computer network administration, auditing, truck driving)

Write in:

16. Could this certification or license be used if you wanted to get a job with any employer in that line of work? If you have a state certification or license that can be used state-wide, please answer "yes".

☐ Yes

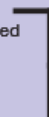
☐ No

17. Is this certification or license for the job you have now? If you are currently not employed, please answer "no". (Mark one.)

☐ Yes, and it is required for my job

☐ Yes, and it is NOT required for my job

☐ No



**GO TO question 20**

18. Is this certification or license for a job you held in the past or for a job you plan to hold in the future? (Mark "Yes" or "No" for each.)

Yes No



For a job that I held in the past..... ☐ ☐

For a job that I plan to hold in the future..... ☐ ☐

19. Other than your most recent certification or license, do you have another certification or license for the job you have now? If you are currently not employed, please answer "no".

☐ Yes

☐ No

► Continue on the next page.



## Educational Certificates

➤ These next questions ask about education you might have received after high school. Include educational certificates you may have earned as part of getting a professional certification or license, but report only information for the educational certificate (not for the certification or license).

- 20. Some people complete a program of study at a vocational or trade school, community or technical college, or other college or university in order to earn an educational certificate rather than a degree. Sometimes this is called a vocational diploma, for example, a cosmetology or mechanics diploma, which differs from a high school diploma. Have you ever earned this type of educational certificate?**

☐ Yes

☐ No → **GO TO question 29**

- 21. Thinking of all the educational certificates you have, which one of the following best describes them? (Mark one.)**

☐ All of my educational certificates were for people who **HAVE** a bachelor's degree

→ **GO TO question 29**

☐ One or more of my educational certificates were for people who **DO NOT HAVE** a bachelor's degree

➤ Of the educational certificates or diplomas you earned after high school, we would like to know a few details about the one you earned most recently. We will use the word "certificate" to refer to this educational certificate or diploma.

➤ Please answer the questions in this section about this **MOST RECENT** certificate.

- 22. In what year did you earn this MOST RECENT certificate?**

Write in year:

- 23. How long did it take you to earn this certificate? (Mark one.)**

☐ Less than 10 weeks (2½ months)

☐ 10 weeks (2½ months) or more, but less than one year

☐ One year or more

☐ I don't know

- 24. What type of school awarded this certificate? (Mark one.)**

☐ Trade, vocational, or business school

☐ Community or technical college

☐ Other college or university

☐ Other type of school (specify) ↴

► **Continue on the next page.**

251 03078

NATES-10AC

**25. Why did you get this certificate? (Mark "Yes" or "No" for each.)**

	Yes ▼	No ▼
To get a job in a new field .....	<input type="checkbox"/>	<input type="checkbox"/>
To get a promotion or raise in pay .....	<input type="checkbox"/>	<input type="checkbox"/>
To stay current in my field or expand skills in my field .....	<input type="checkbox"/>	<input type="checkbox"/>
To start my own business. ....	<input type="checkbox"/>	<input type="checkbox"/>
To get a professional certification or license. ....	<input type="checkbox"/>	<input type="checkbox"/>
Other (specify) ↴ .....	<input type="checkbox"/>	<input type="checkbox"/>

**26. What was the primary subject or field of study for this certificate?**

Write in:

**27. Is the subject field of this certificate related to the job you have now? If you are currently not employed, please answer "no."**

☐ Yes → **GO TO question 29**

☐ No

**28. Is the subject field of this certificate related to a job you held in the past or to a job you plan to hold in the future? (Mark "Yes" or "No" for each.)**

	Yes ▼	No ▼
Related to a job that I held in the past. ....	<input type="checkbox"/>	<input type="checkbox"/>
Related to a job that I plan to hold in the future. ....	<input type="checkbox"/>	<input type="checkbox"/>

► **Continue on the next page.**



## Apprenticeships

29. In a formal apprenticeship program, an apprentice receives both instruction and on-the-job training and is paid a training salary. Have you ever participated in this type of apprenticeship program? (Do not count student teaching, medical internship or residency, or unpaid internships. Mark one.)

- ☐ Yes, I have COMPLETED this type of program
- ☐ Yes, I am currently participating in this type of program
- ☐ No, I have not participated in this type of program

**GO TO  
question  
35**

30. In what year did you complete this apprenticeship program?

Write in year:

31. Did this apprenticeship program lead to a Certificate of Completion of Apprenticeship from your state or from the U.S. Department of Labor?

- ☐ Yes
- ☐ No
- ☐ I don't know

32. As part of this apprenticeship program, did you take any courses from a community or technical college or from another college or university?

- ☐ Yes
- ☐ No

33. Was this apprenticeship program for the job or industry you work in now? If you are currently not employed, please answer "no".

- ☐ Yes → **GO TO question 35**
- ☐ No

34. What occupation was this apprenticeship program for? (for example: carpenter, electrician, water treatment operator, emergency medical technician)

Write in:

► Continue on the next page.

251 03094

NATES-10AC

## College Classes

35. Are you currently taking classes from a vocational or trade school, community or technical college, or other college or university? If you are on spring, summer, or holiday break, please answer "yes".

☐ Yes

☐ No → **GO TO question 46**

➤ The rest of this section asks about these college classes. If you are on a school break, please respond for the classes you were taking before you went on break.

36. Are you taking these classes to earn a diploma, certificate, or degree? (Do not count professional certifications or licenses.)

☐ Yes

☐ No → **GO TO question 39**

37. What diploma, certificate, or degree are you earning? (Mark one.)

☐ Diploma or certificate below the bachelor's degree level

☐ Associate's degree (for example, AA, AS, AAS)

☐ Bachelor's degree (for example, BA, AB, BS, BFA)

☐ Certificate above the bachelor's degree level

☐ Master's degree (for example, MA, MS, MEng, MEd)

☐ Professional or doctorate degree (for example, MD, DDS, DVM, LLB, JD, PhD, EdD)

38. Are you going to school full time or part time? (Mark one.)

☐ Full time

☐ Part time

→ **GO TO question 46**

39. How many classes are you currently taking? (Mark one.)

☐ One class

☐ Two or more classes

40. Which ONE of the following best describes the type of classes you are taking? (Mark one.)

☐ All my classes are for college credit

☐ Some of my classes are for college credit, some are not for credit

☐ All my classes are not for credit

☐ I don't know whether my classes are for college credit

41. Why are you taking these classes? (Mark "Yes" or "No" for each.)

	Yes ▼	No ▼
To get a job in a new field. ....	<input type="checkbox"/>	<input type="checkbox"/>
To get a promotion or raise in pay .....	<input type="checkbox"/>	<input type="checkbox"/>
To stay current in my field or expand skills in my field .....	<input type="checkbox"/>	<input type="checkbox"/>
To start my own business .....	<input type="checkbox"/>	<input type="checkbox"/>
To get a professional certification or license .....	<input type="checkbox"/>	<input type="checkbox"/>
To earn continuing education or other professional development credits .....	<input type="checkbox"/>	<input type="checkbox"/>
To help me decide if I want to get a diploma, certificate, or degree. ....	<input type="checkbox"/>	<input type="checkbox"/>
Classes are required to enter a college program .....	<input type="checkbox"/>	<input type="checkbox"/>
Personal interest in the subject of the classes. ....	<input type="checkbox"/>	<input type="checkbox"/>
Other (specify) ↴ .....	<input type="checkbox"/>	<input type="checkbox"/>

42. What is the primary subject or field of study for these classes? If you are taking classes in different subjects, please list the subject you consider your main interest or focus.

Write in:

▶ **Continue on the next page.**

251 03102

**43. Did your employer require that you take any of these classes?**

☐ Not relevant – I am not employed or I am self-employed



**GO TO question 46**

☐ Yes

☐ No

**44. For any of these classes, is your employer paying your tuition or fees, or reimbursing you for your tuition or fees? (Mark one.)**

☐ Yes, my employer is paying all of the tuition and fees

☐ Yes, my employer is paying part of the tuition and fees

☐ No

**45. Are any of these classes designed specifically for employees at your company?**

☐ Yes

☐ No

☐ I don't know

► **Continue on the next page.**



## Other Instruction or Training

- 46. OTHER THAN APPRENTICESHIPS AND COLLEGE CLASSES YOU MAY HAVE DESCRIBED EARLIER, in the past 12 months, have you completed any other courses, training, or formal instruction, either at work or outside of work?**

*This includes both work or personal interest courses, seminars, webinars, or workshops on topics such as:*

- job safety, work ethics or other regulations
- equipment use
- communication, sensitivity, or team-building
- computer or technical skills
- management skills
- other job skills
- fitness classes, art, dance, or music lessons, religious education
- learning to speak English
- basic skills education classes
- other topics not listed here

**Have you completed any such instruction or training IN THE PAST 12 MONTHS?**

☐ Yes

☐ No → **GO TO question 52**

- 47. In the past 12 months, which of the following types of instruction or training have you completed, either online or in-person? (Mark all that apply.)**

### Job training

- ☐ SAFETY AND COMPLIANCE TRAINING (includes information on company or professional procedures and regulations concerning legal, ethical, and safety issues)
- ☐ COMMUNICATION, SENSITIVITY, OR TEAM TRAINING (includes training to improve communication in the workplace, encourage teamwork, or to reorganize work teams and work flow)
- ☐ MANAGEMENT TRAINING (includes training in supervising employees and in implementing employment practices, regulations, and policies)
- ☐ JOB SKILLS TRAINING (includes training to develop the skills you need to do your work, such as sales and customer relations training, professional or technical skill development, use of computer applications, and other practical job skills)

### Basic skills education

- ☐ BASIC READING, WRITING, OR ARITHMETIC INSTRUCTION (instruction for adults below the high school level)
- ☐ HIGH SCHOOL COMPLETION (classes to prepare for the GED or other adult high school program)
- ☐ ENGLISH LANGUAGE INSTRUCTION (classes to learn to speak English)

### Other instruction or training

- ☐ PERSONAL INTEREST OR DEVELOPMENT (instruction related to hobbies and interests outside of work)
- ☐ OTHER (specify) →

► **Continue on the next page.**

251 03128

NATES-10AC

**48. Did any ONE of these activities last at least 8 hours (either in one session or across multiple sessions)?**

☐ Yes

☐ No → **GO TO question 52**

➤ In the questions below, we ask a few details about your most recently completed instruction or training that lasted at least 8 hours.

➤ Please answer these questions only about this MOST RECENT completed instruction or training that lasted at least 8 hours.

**49. Which ONE of the following best describes this MOST RECENT instruction or training? (Mark one.)**

☐ Safety and compliance training

☐ Communication, sensitivity, or team training

☐ Management training

☐ Job skills training

☐ Basic reading, writing, or arithmetic instruction

☐ High school completion

☐ English language instruction

☐ Personal interest or development

→ **GO TO question 52**

☐ Other (specify) ↴

**50. Was this instruction or training that your employer offered at no charge during working hours?**

☐ Yes

☐ No

☐ Not relevant – I was self-employed or not employed when I took the instruction or training

**51. Why did you take this instruction or training? (Mark "Yes" or "No" for each.)**

	Yes ▼	No ▼
To get a job in a new field . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To get a promotion or raise in pay . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To stay current in my field or expand skills in my field . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To start my own business. . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To get a professional certification or license . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To earn continuing education or other professional development credits . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
To meet an employer requirement . . . . .	<input type="checkbox"/>	<input type="checkbox"/>
Other (specify) ↴ . . . . .	<input type="checkbox"/>	<input type="checkbox"/>

► **Continue on the next page.**





## Employment

**52. LAST WEEK, were you employed for pay at a job or business, or were you temporarily absent from a job or business?**

☐ Yes – was working (even for as little as one hour) or was temporarily absent from work (on vacation, temporary illness, maternity leave, other family/personal reasons, bad weather, etc.)

☐ No – was not employed, was on layoff, or was retired

GO TO question 56

**53. For the job or business you were in last week, were you a member of a labor union or of an employee association similar to a union? (for example, AFL-CIO, Change to Win Federation, NEA)**

☐ Yes

☐ No

**54. Which one of the following best describes your employment situation last week (or when you last worked)? (Mark one.)**

☐ I worked at a full-time job (job of 35 hours or more per week)

GO TO question 61

☐ I worked at one or more part-time jobs (no full-time job)

**55. Would you have preferred to work at a full-time job?**

☐ Yes

☐ No

GO TO question 61

**56. LAST WEEK, were you on layoff from a job?**

☐ Yes

☐ No

**57. During the LAST 4 WEEKS, have you been ACTIVELY looking for work?**

☐ Yes

☐ No

GO TO question 59

**58. LAST WEEK, could you have started a job if offered one, or returned to work if recalled?**

☐ Yes, I could have gone to work

☐ No, because of my own temporary illness

☐ No, because of some other reason (in school, etc.)

GO TO question 60

**59. Do you intend to look for work within the next 5 years?**

☐ Yes

☐ No

☐ I don't know

**60. When did you last work, even for a few days?**

☐ Within the past 12 months

☐ Over 12 months ago

GO TO question 63

☐ Never worked for pay

GO TO question 70

**61. During the PAST 12 MONTHS (52 weeks), how many weeks did you work, even for a few hours, INCLUDING paid vacation, paid sick leave, and military service?**

☐ 50 to 52 weeks

☐ 48 to 49 weeks

☐ 40 to 47 weeks

☐ 27 to 39 weeks

☐ 14 to 26 weeks

☐ 13 weeks or less

**62. During the PAST 12 MONTHS, in the WEEKS WORKED, how many hours did you usually work each WEEK?**

Usual hours worked each WEEK

GO TO question 64

► Continue on the next page.

251 03144

NATES-10AC

**63. Please write in the month and year for when you last worked.**

--	--	--	--	--	--	--	--

month                      year



**GO TO question 65**

**64. Which category best fits your earnings from wages, salary, commissions, bonuses or tips, from all jobs over the PAST 12 MONTHS? Report amount before deductions for taxes, bonds, dues or other items. (Mark one.)**

- ☐ \$0 to \$10,000
- ☐ \$10,001 to \$20,000
- ☐ \$20,001 to \$30,000
- ☐ \$30,001 to \$40,000
- ☐ \$40,001 to \$50,000
- ☐ \$50,001 to \$60,000
- ☐ \$60,001 to \$75,000
- ☐ \$75,001 to \$150,000
- ☐ \$150,001 or more

**65. For the next few questions, please describe clearly your chief job activity or business last week. If you had more than one job, describe the one at which you worked the most hours. If you had no job or business last week, give information for your last job or business.**

**In your current or last job, which ONE of the following were you? (Mark one.)**

- ☐ An employee of a private for-profit company or business, or of an individual, for wages, salary, or commissions
- ☐ An employee of a private not for-profit, tax exempt, or charitable organization
- ☐ A local government employee (city, county, etc.)
- ☐ A state government employee
- ☐ A Federal government employee
- ☐ Self-employed in own business, professional practice, or farm
- ☐ Working without pay in family business or farm

**66. For whom did you work?**

- ☐ If now on active duty in the Armed Forces, mark (X) this box and print the branch of the Armed Forces below.

Name of company, business, or other employer

**67. What kind of business or industry was this?**

*Describe the activity at the location where employed. (for example: hospital, newspaper publishing, mail order house, auto engine manufacturing, bank)*

**68. What kind of work were you doing?**

*(for example: registered nurse, personnel manager, supervisor of order department, secretary, accountant)*

**69. What were your most important activities or duties?**

*(for example: patient care, directing hiring policies, supervising order clerks, typing and filing, reconciling financial records)*

► **Continue on the next page.**



## Background

70. Are you male or female?

- ☐ Male  
☐ Female

71. How old are you?

years old

72. What is your current marital status?

- ☐ Now married → **GO TO question 75**  
☐ Widowed  
☐ Divorced  
☐ Separated  
☐ Never married

73. Are you currently living with a boyfriend/girlfriend or partner in this household?

- ☐ Yes  
☐ No → **GO TO question 75**

74. Are you currently in a registered domestic partnership or civil union?

- ☐ Yes  
☐ No

75. Are you of Hispanic or Latino origin?

- ☐ Yes  
☐ No

76. What is your race? Choose one or more.

- ☐ White  
☐ Black or African American  
☐ Asian  
☐ American Indian or Alaska Native  
☐ Native Hawaiian or other Pacific Islander

77. Do you speak a language other than English at home?

- ☐ Yes  
☐ No → **GO TO question 79**

78. How well do you speak English? (Mark one.)

- ☐ Very well  
☐ Well  
☐ Not very well  
☐ Not at all

79. Where were you born? (Mark one.)

- ☐ In the United States (the 50 states or the District of Columbia)  
→ **GO TO question 82**  
☐ In a U.S. territory (Puerto Rico, Guam, American Samoa, U.S. Virgin Islands, or Northern Marianas)  
→ **GO TO question 82**  
☐ Outside the U.S. (in a foreign country)

80. Were you born abroad to one or more parents who were U.S. citizens?

- ☐ Yes  
☐ No

81. When did you come to live in the United States?

Year

► Continue on the next page.

251 03169

**82. Which category best fits the total income of all persons in your household over the past 12 months? Include your own income. Include money from jobs or other earnings, pensions, interest, rent, Social Security payments, and so on. (Mark one.)**

- ☐ \$0 to \$10,000
- ☐ \$10,001 to \$20,000
- ☐ \$20,001 to \$30,000
- ☐ \$30,001 to \$40,000
- ☐ \$40,001 to \$50,000
- ☐ \$50,001 to \$60,000
- ☐ \$60,001 to \$75,000
- ☐ \$75,001 to \$150,000
- ☐ \$150,001 or more

### **Thank you.**

*Please return this questionnaire in the postage-paid envelope provided. If you need additional questionnaires for other eligible household members, please call the Census Bureau toll-free at 1-800-845-8243.*

*If you have lost the envelope, mail the completed questionnaire to:*

**U.S. Census Bureau  
ATTN: DSB 60-A  
1201 E. 10th Street  
Jeffersonville, IN 47132-0001**

251 03177



## Appendix G: Nonresponse Follow-up Study (NRFU) Questionnaire

FORM: **NRBS-1A**  
(03/21/2013)

OMB No. 1850-0803: Approval Expires 09/30/2013  
Conducted by:  
U.S. DEPARTMENT OF COMMERCE  
Economics and Statistics Administration  
U.S. CENSUS BUREAU

## 2013 National Adult Training and Education Survey

### SECTION 1

1. **FR Name:**

2. **FR Code:**

3. **RO Code:**

4. **Address:**

5. **Interview Date:**

Month Day Year

--	--	--

6. **Contact Attempts:**

- ☐ 0
- ☐ 1-5
- ☐ 6-10
- ☐ 11 +

7. **Interview Mode:**

(skip this section if a LMR)

- ☐ Personal visit
- ☐ Transcribed at headquarters
- ☐ Telephone (only use if no other way to get information)

8. **Outcome Codes:**

Complete

- ☐ Interviews - 201
- ☐ Mailout Questionnaire Received - 205

Type As

- ☐ Language Barrier - 213
- ☐ Refusal - 218

Type Bs

- ☐ Temporarily Absent - 217
- ☐ Vacant - 226

Type Cs

- ☐ Multi Unit Address Without Unit Designation - 247
- ☐ Out of Scope - 248
- ☐ Unable to Locate/Bad Address - 249
- ☐ Late Mail Return - 250



## SECTION 2

**A.** Hello, my name is . I work for the U.S. Census Bureau. We are conducting a short 5-minute survey asking about adults' education and training. For this survey, I need to speak to an adult between the ages of 16 and 65 who is no longer in high school. Do you or someone in this household fit this description?

- ☐ Yes, person answering the door fits this description → Go to INTRO TEXT
- ☐ Yes, but not the person answering the door → Ask to speak to someone who fits that description, then go to A.
- ☐ No → Thank them for their time and say there are no further questions.

INTRO TEXT: Great. First, you need to know that your participation in this survey is voluntary and all information you give is confidential. Your responses may be used only for statistical purposes and may not be disclosed or used in identifiable form for any other purpose except as required by law. I'm going to start with a couple of questions about your household.



## Survey Starts Here

1. Including you, how many of the people living in this house are ages 16 to 65?

2. Including you, how many of these people ages 16 to 65 are no longer in high school?

3. Now I have some questions about you. Looking at this card, what is the highest degree or level of school you have completed? *[Show flashcard 1]*

### FLASHCARD 1

- ☐ Elementary or high school, but no high school diploma or GED
- ☐ High school diploma, GED, or other high school completion
- ☐ Some college credit but less than one year of college credit
- ☐ 1 or more years of college credit, no degree
- ☐ Associate's degree (for example, AA, AS)
- ☐ Bachelor's degree (for example, BA, BS)
- ☐ Master's degree (for example, MA, MS, MEng, MEd, MSW, MBA)
- ☐ Professional degree beyond a bachelor's degree (for example, MD, DDS, DVM, LLB, JD)
- ☐ Doctorate degree (for example, PhD, EdD)
- ☐ I don't know
- ☐ Refused

4. Do you have a professional certification or a state or industry license? A professional certification or license shows you are qualified to perform a specific job and includes things like licensed realtor, certified medical assistant, certified construction manager, a project management professional certification, or an IT certification.

- ☐ Yes
- ☐ No
- ☐ I don't know
- ☐ Refused

GO TO question 6

5. Is your certification or license for a job in teaching, in health care, in accounting, finance or insurance, or in some other field? If you have more than one certification or license, please answer for the one that you earned most recently. (Mark One)

- ☐ Teaching
- ☐ Health care
- ☐ Accounting, finance, or insurance
- ☐ Other
- ☐ I don't know
- ☐ Refused

6. Some people complete a program of study at a vocational or trade school, community or technical college, or other college or university in order to earn an educational certificate rather than a degree. Sometimes this is called a vocational diploma, for example, a cosmetology or mechanics diploma, which differs from a high school diploma. Have you ever earned this type of educational certificate?

- ☐ Yes
- ☐ No
- ☐ I don't know
- ☐ Refused

► Continue on the next page.





- 7. Are you currently taking classes from a vocational or trade school, community or technical college, or other college or university? If you are on spring, summer, or holiday break, please answer "yes".**

☐ Yes  
☐ No  
☐ I don't know  
☐ Refused

- 8. I now have just a few background questions. How old are you?**

Age in years

☐ Refused

- 9. Are you of Hispanic or Latino origin?**

☐ Yes  
☐ No  
☐ Refused

- 10. Looking at this card, what is your race? Choose one or more.**  
*[Show flashcard 2]*

**FLASHCARD 2**

☐ White  
☐ Black or African American  
☐ Asian  
☐ American Indian or Alaska Native  
☐ Native Hawaiian or other Pacific Islander  
☐ Refused

- 11. *[If not apparent]* What is your sex?**

☐ Male  
☐ Female  
☐ Refused

- 12. I have two more questions about your household. Looking at this card, which category best fits the total income of all persons in your household over the past 12 months? *[Show flashcard 3]***

*Probe: Include your own income. Include money from jobs or other earnings, pensions, interest, rent, Social Security payments, and so on.*

**FLASHCARD 3**

☐ \$0 to \$10,000  
☐ \$10,001 to \$20,000  
☐ \$20,001 to \$30,000  
☐ \$30,001 to \$40,000  
☐ \$40,001 to \$50,000  
☐ \$50,001 to \$60,000  
☐ \$60,001 to \$75,000  
☐ \$75,001 to \$150,000  
☐ \$150,001 or more  
☐ I don't know  
☐ Refused

- 13. The Census Bureau mailed a survey package to your house a few weeks ago. Did your household receive that package?**

☐ Yes  
☐ No  
☐ I don't know  
☐ Refused

**Those are all the questions I have.**

**Thank you for your time.**

