UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

November 15, 2023

2023 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS23-RER-10

MEMORANDUM FOR Donna Daily
Chief, American Community Survey Office

From: David Waddington *David Waddington*
Chief, Social, Economic, and Housing Statistics Division

Prepared by: Kevin McElrath *K.M.*, Jacob Fabina *J.S.F.*, Michael J. R. Martin *M.J.R.M.*
Social, Economic, and Housing Statistics Division
Broderick E. Oliver *B.O.*
Decennial Statistical Studies Division
Joseph P. Gilmartin *J.G.*
American Community Survey Office

Subject: 2022 American Community Survey Content Test Evaluation Report: Educational Attainment

Attached is the 2022 American Community Survey (ACS) Content Test report for Educational Attainment. This report presents the methods and results of the test for a revised version of the educational attainment question.

If you have any questions about this report, please contact Kevin McElrath at 301-763-6175.

cc:
acso.re.workgroup.list@census.gov

Lauren Contard	DSSD	Stephanie Galvin	SEHSD
Lindsay Longsine	DSSD	Erik Hernandez	SEHSD
Peter Massarone Jr.	DSSD	Hyon Shin	SEHSD
Michael Risley	DSSD		
Samantha Spiers	DSSD		

This page is intentionally blank.

DRB Clearance Number—CBDRB-FY23-ACSO003-B0062

American Community Survey Research and Evaluation Program

November 15, 2023

2022 American Community Survey Content Test Evaluation Report: Educational Attainment

FINAL REPORT

Kevin McElrath
Jacob Fabina
Michael J. R. Martin
Social, Economic, and Housing Statistics Division
Broderick E. Oliver
Decennial Statistical Studies Division
Joseph P. Gilmartin
American Community Survey Office

This page is intentionally blank.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The U.S. Census Bureau conducted the 2022 American Community Survey (ACS) Content Test, from September through December of 2022. The 2022 ACS Content Test tested the wording, format, and placement of proposed new ACS questions and proposed revisions of current ACS questions for potential inclusion in the ACS data collection instruments. The tested questions came from 10 topics. This report presents the results of this field test for Educational Attainment.

In preparation for the 2022 Content Test, the Census Bureau, in consultation with the Office of Management and Budget (OMB) and the Interagency Council on Statistical Policy Subcommittee on the ACS, determined which proposals solicited from over 25 federal agencies would be tested in 2022. Approved proposals for new content or changes to existing content were tested according to the ACS content change process, which includes cognitive testing and field testing.

The 2022 ACS Content Test consisted of a nationally representative sample of 120,000 housing unit addresses, excluding Puerto Rico, Alaska, and Hawaii. The sample, which was independent of production ACS, was divided evenly among three treatments, a Control treatment and two test treatments.

Like production ACS, the data collection for the 2022 ACS Content Test was conducted in two phases: a self-response phase, which lasted up to nine weeks, followed by a nonresponse follow-up phase, conducted via Computer-Assisted Personal Interviewing (CAPI). The CAPI operation lasted about one month. For Households where we received a response in the original Content Test interview, a Content Follow-Up telephone reinterview was conducted to measure response error.

Since 2008, headers were added to each section of the educational attainment section, including “No schooling completed.” With the addition of these headers, the rate of respondents selecting “No schooling completed” increased and may have erroneously included adults who completed some schooling.

The purpose of this field test was to determine if changes to the response options of the educational attainment question would reduce the number of cases where respondents who have completed some schooling, select the “No schooling completed” response category. The Control version of the question has “No schooling completed” as the lowest response category, followed by two additional response categories for educational attainment below grade 1: “Nursery school” and “Kindergarten.” The Test version collapses these three response categories into one new category: “Less than grade 1.”

The results presented in this report support recommending the Test version of the educational attainment question for implementation in the ACS. The percentage of adults who report their educational attainment is “Less than grade 1” is significantly lower in the Test version than in the Control version. The recommendation, based on this research, is to implement the Test version of the educational attainment question.

1 BACKGROUND

The U.S. Census Bureau conducted the 2022 American Community Survey (ACS) Content Test from September to December of 2022. The 2022 ACS Content Test tested the wording, format, and placement of proposed new ACS questions and proposed revisions of current ACS questions for potential inclusion in the ACS data collection instruments. The questions came from these ten ACS topics, three of which, Sewer, Electric Vehicles, and Solar Panels are new:

- Household Roster
- Sewer
- Electric Vehicles
- Solar Panels
- Supplemental Nutrition Assistance Program (SNAP)
- Educational Attainment
- Health Insurance Coverage
- Disability
- Labor Force
- Income

This report presents the results of the field test for Educational Attainment.

1.1 Proposals for New and Revised ACS Questions

In June 2018, the Census Bureau solicited proposals for new or revised ACS content from over 25 federal agencies. For new questions, the proposals explained why these data were needed and why other data sources that provide similar information were not sufficient. Proposals for new content were reviewed to ensure that the requests met a statutory or regulatory need for data at small geographic levels or for small populations.

The Census Bureau, in consultation with the Office of Management and Budget (OMB) and the Interagency Council on Statistical Policy Subcommittee on the ACS, determined which proposals moved forward. Approved proposals for new content or changes to current content were tested via the ACS content change process. This process includes cognitive testing and field testing. An interagency team consisting of Census Bureau staff and representatives from other federal agencies participated in development and testing activities.

In accordance with OMB's Standards and Guidelines for Statistical Surveys (OMB, 2006) and the Census Bureau's Statistical Quality Standards (U.S. Census Bureau, 2022a), the Census Bureau conducted cognitive interviewing to pretest survey questions prior to field testing or implementing the questions in production.

1.2 Cognitive Testing

For the 2022 ACS Content Test, the Census Bureau contracted with Research Triangle Institute (RTI) International to conduct three rounds of cognitive testing.[1] Cognitive interviews were conducted virtually, in English and Spanish.[2] In the first round of cognitive testing, each topic tested one or two versions of the question. Based on the results of the first round, wording modifications to the questions were made and one or two versions per topic were tested in the second round. The interagency team used the results of both rounds of cognitive testing to recommend question content for the field test. For more information on the cognitive testing procedures and results from rounds one and two, see RTI International (2022a).

The third round of cognitive testing was conducted in Puerto Rico and in Group Quarters (GQ), as the 2022 ACS Content Test did not include field testing in these areas. Cognitive interviews in Puerto Rico were conducted in Spanish; GQ cognitive interviews were conducted in English. For more information on the cognitive testing procedures and results from the third round, see RTI International (2022b).

Three topics included in the cognitive testing were not included in the field test: Homeowners Association or Condominium Fees, Home Heating Fuel, and Means of Transportation to Work. For the most part, the changes to these questions are expected to either impact a small population or result in a small change in the data that would not be detectable in the Content Test. The subject matter experts recommended that cognitive testing was sufficient for these questions and that field testing was not necessary; the Interagency Council on Statistical Policy Subcommittee on the ACS agreed with this recommendation. Content changes for these topics will be implemented in production ACS in 2024.

1.3 Field Testing Educational Attainment in the 2022 ACS Content Test

1.3.1 Justification for Inclusion of Educational Attainment in the Content Test

A question on educational attainment has been asked on the ACS since the survey began in 2005. The current version of the question has been asked since 2008. Ongoing research at the Census Bureau shows that a relatively high percentage of people are selecting the response category “No schooling completed” in the self-response modes of the ACS. This testing evaluates whether changes to the lowest response categories can reduce the number of cases

[1] For each test topic, subcommittees were formed to develop question wording and research requirements for cognitive testing. The subcommittees included representation from the Census Bureau and other federal agencies.

[2] Cognitive testing interviews were conducted virtually due to the COVID-19 pandemic. Interviews were attempted by videoconferencing first and were moved to phone interviews if there were technical problems with Skype or MS Teams.

where respondents with some schooling completed, select the "No schooling completed" response category.

Starting in 2008, headers were added to the educational attainment question to differentiate between various levels of attainment.[3] These headers included "No schooling completed", which was added as a header during the 2006 ACS Content Test after cognitive testing suggested that respondents had difficulty locating the "No schooling completed" category (Crissey, Bauman, and Peterson, 2007). The Educational Attainment report for the 2006 ACS Content Test noted a higher rate of "No schooling completed" in the revised educational attainment question.

In 2007, 0.71 percent of persons 25 years and older who responded by mail, selected "No schooling completed." In 2008, following the revision, 1.52 percent of persons 25 years and older who responded by mail selected "No schooling completed." In 2013, the ACS introduced an internet response option. An analysis of 2016 data found that 1.22 percent of persons 25 years and older who responded by internet and 2.26 of persons who responded by mail selected "No schooling completed."

The percent of persons 25 years and older who selected "No schooling completed" in the Computer Assisted Telephone Interview (CATI) and Computer Assisted Personal Interview (CAPI) response modes were 0.68 and 0.88 percent, respectively, in 2007; 0.66 and 0.83, respectively, in 2008; and 0.65 and 0.85, respectively, in 2016. Some of these differences between 2008 and 2016 may be attributable to changing demographics of those who respond in each mode, which can be isolated during a content test. Immigrants are likely to have different levels of educational attainment than native born persons, as immigrants often immigrate to the U.S. to fill specific high- or low-skill labor niches (Lee and Zhou, 2015). Hence, the foreign-born population is an important subpopulation to test the lower response categories in the educational attainment question.

1.3.2 Cognitive Testing Development for Educational Attainment

Cognitive testing was performed on multiple modes across several rounds to identify a final test version for the Content Test. The major focus of the cognitive testing was three-fold. First, we sought to ascertain how the headers on the educational attainment question impact respondents' selection of the appropriate category. Second, we examined if there were differences in the way respondents understood "grade" versus "level" of schooling. Third, we assessed how collapsing categories for the lowest levels of educational attainment impacted respondents' ability to select the correct educational attainment category.[4]

[3] See the control instrument on page 6 of this report to view the headers.

[4] For more information and to see the wording and modes tested, see RTI International, 2022a and 2022b.

The key research questions and their associated results of cognitive testing are as follows:

1. Do participants understand the question and the overall meaning?

   Results: Overall, participants understood the question as intended and answered accurately. Only three of the 33 participants answered the educational attainment question incorrectly—two of these participants answered incorrectly for other household members and one answered incorrectly for themselves (RTI International 2022a; 2022b).

2. Is the use of the term "grade" (version 1) or "level" (version 2) of school in the base question easier to understand?

   Results: Participants interpreted both phrases in the same way – that is, "highest grade of school" and "highest level of school" were interpreted to mean the highest education one had received. However, when asked about the terms "grade" and "level" separately, participants shared a variety of interpretations. Despite various definitions of "grade" vs. "school," participants used these terms interchangeably during probing, and the terminology used in the version they received did not seem to affect their response. Overall, it seems as if, within the context of the educational attainment question, participants understand these phrases in the same way (RTI International 2022a; 2022b).

3. How do people currently enrolled in school answer this question? Does the instruction help them answer?

   Results: Only one participant currently enrolled in school answered this question incorrectly. This participant answered "Regular high school diploma" then explained during probing that they had completed some school before dropping out of high school, and they were currently working on their GED. All other participants currently enrolled in school answered correctly—that is, answered for the highest level of school they had completed. Those currently enrolled in school indicated that they did not notice or read the instructions about how to answer if currently enrolled. These participants did not feel they needed to read these instructions as they already knew the answer, or they assumed all necessary information would be included in the bold text. Though nearly all participants currently enrolled in school answered correctly for themselves, several participants incorrectly reported the highest grade completed when reporting for a young child. This is likely because participants do not seem to be reading the italicized instructions (RTI International 2022a; 2022b).

4. Use of “Less than grade 1” (cognitive testing version 1) vs. expanded three categories “less than 1
   year of school completed,” “nursery school or preschool,” and “kindergarten” (cognitive testing version 2) – Does one version better help people understand who should select these response options?

   Results: Version 2’s expanded categories of “Less than 1 year of school completed,” “Nursery school or preschool,” and “Kindergarten” caused a lot of confusion among participants. Participants did not have a strong understanding of the difference between these categories, and participants provided highly varied definitions of “Less than 1 year of school completed.” When asked about “Less than grade 1,” participants indicated that this response option would include anybody who never went to school or children who had only attended preschool, pre-K, or Kindergarten (RTI International 2022a; 2022b).

5. Unlike the current production version, both test versions in cognitive testing remove the heading and subheading “No schooling completed.” Do those with no schooling know how to respond?

   Results: RTI/RSS (Research Support Services) did not interview any participants with no schooling; however, probing included asking participants how they would answer each version of the question for an adult who had never attended school. Overall, both versions performed similarly, with most participants selecting “Less than grade 1” or “Less than 1 year of school completed” for each version; however, some participants felt an additional option of “Not Applicable” or “Never attended school” should be added for this scenario (RTI International 2022a; 2022b).

6. Do individuals with homeschooled children understand the question and response categories and know where their child should be classified?

   All parents of homeschooled children were able to easily answer the question and knew the grade/level of school in which their child should be classified (RTI International 2022a; 2022b).

   Based on the three rounds of cognitive testing, the committee decided on testing a single alternative for the content test that removed the response category headers and combined the lowest level of schooling into one category as described in the next section.

1.3.3 Question Content

The Control and Test versions of Educational Attainment for the paper data collection instrument are shown in Figure 1 and Figure 2, respectively. The Control version of the question has "No schooling completed" as the lowest response category, followed by two additional response categories for educational attainment below grade 1: "Nursery school" and "Kindergarten." The Test version collapses these three response categories into a new category, "Less than grade 1."

The automated versions (internet and CAPI) have the same content, formatted accordingly.

Figure 1. Control: Paper Version Educational Attainment

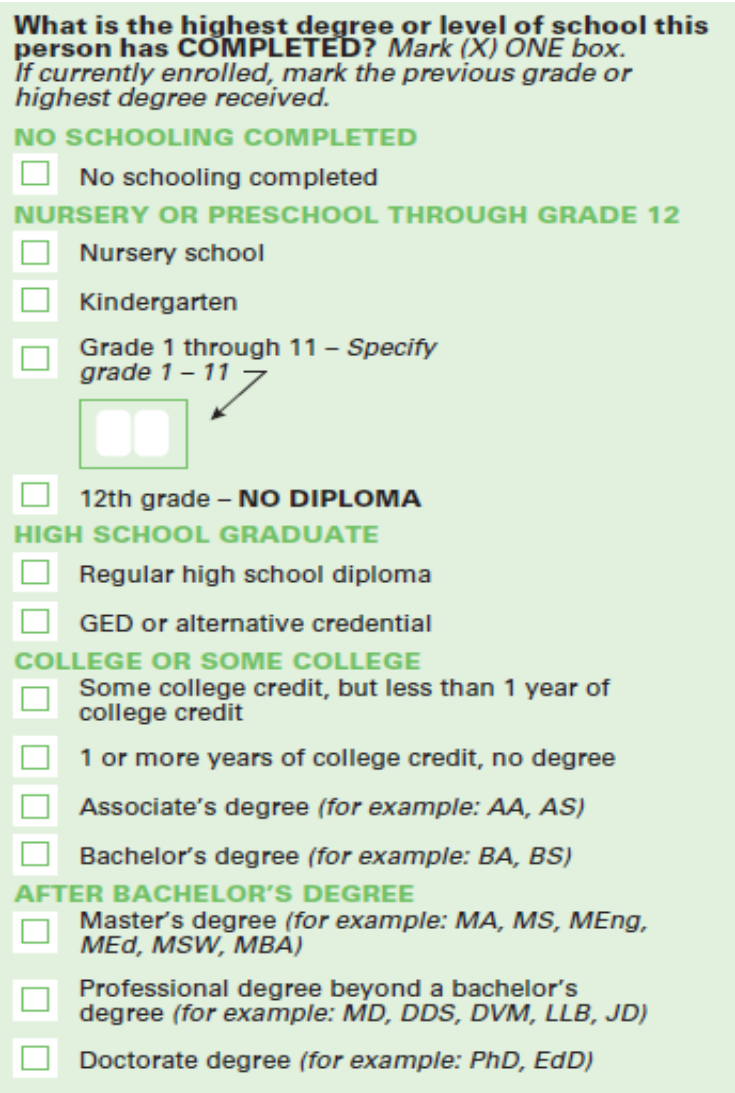
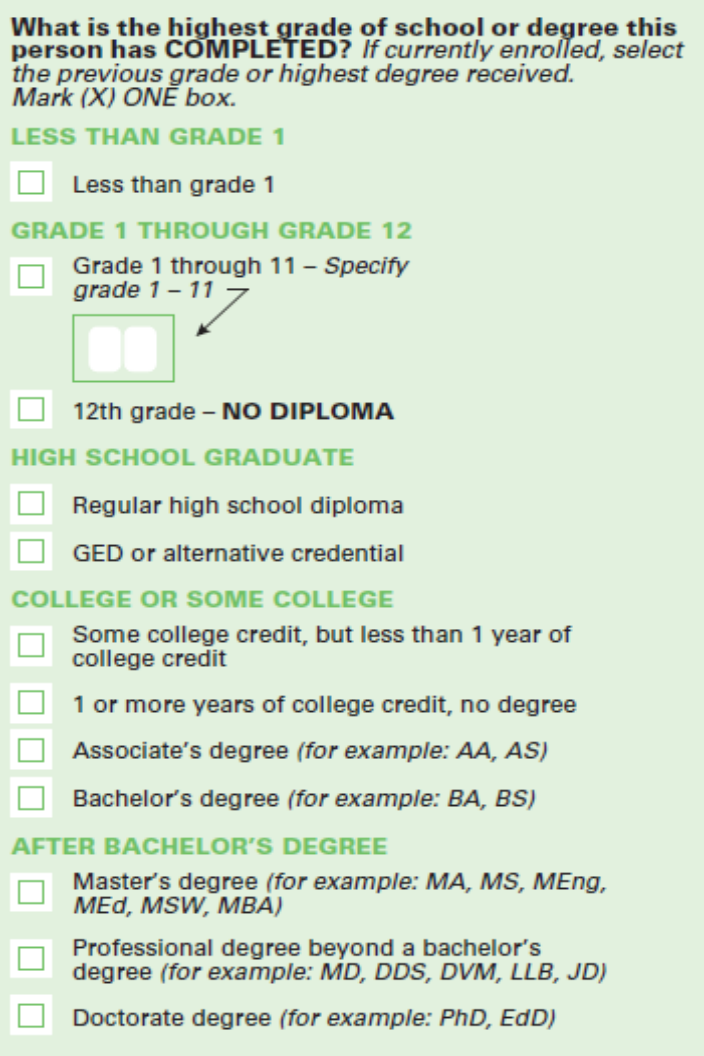


Figure 2. Test: Paper Version Educational Attainment



1.3.4 Research Questions

The questions examined for this research are presented below.

1. *How do the estimates of the percentage of the population with educational attainment "Less than grade 1" in the Control and Test treatments compare with the estimate in the Current Population Survey Annual Social and Economic Supplement (CPS ASEC)?*

2. *Is the missing data rate different for the Test treatment than for the Control treatment?*

3. *Are these distributions of educational attainment for the Test and Control treatments different?*

    - Less than grade 9
    - Grades 9 through 12 (no diploma)
    - High school graduate (regular high school diploma, GED or alternative credential)
    - Some college credit, no degree
    - Associate's degree
    - Bachelor's degree or higher

4. *Are the measures of response reliability (Gross Difference Rate, Index of Inconsistency) for the Control and Test treatments different?*

2 METHODOLOGY

2.1 Sample Design

The 2022 ACS Content Test consisted of a national sample of roughly 120,000 housing unit addresses, excluding Puerto Rico, Alaska, and Hawaii (due to cost constraints, only stateside housing units were included). The sample was independent of the ACS production sample; however, the sample design for the Content Test was largely based on the ACS production sample design, with some modifications to meet the test objectives. The ACS production sample design is described in Chapter 4 of the ACS and Puerto Rico Community Survey (PRCS) Design and Methodology report (U.S. Census Bureau, 2022b).

The sample design modifications included stratifying addresses into high and low self-response areas, oversampling addresses from the low self-response areas to ensure equal response from both strata, and selecting an initial sample of addresses, followed by a nearest neighbor method for selecting the remaining addresses for sample. The high and low self-response strata were defined based on ACS self-response rates from the 2018 and 2019 panels at the tract level.

In the sample selection process, we selected an initial sample of 40,000 addresses, then selected the two nearest neighbors for each initially selected address. If possible, we selected nearest neighbors that were in both the same content test sampling stratum as well as the same state, county, and sub-county area as the initially selected address. In total, three samples were selected, one for the Control treatment and two for the two test treatments. These three treatments are shown in Table 1.

The Control treatment contained production questions and questions from the three new topics: Solar Panels, Electric Vehicles, and Sewer. The Test treatment contained a test version

question for all topics except Household Roster. Two of the new topics, Solar Panels and Sewer, only had one version of the test question; therefore, the same question was asked in the Control and test treatments. The other new topic, Electric Vehicles, had two versions; one was asked in the Control and Roster Test treatments and the other in the Test treatment.

The primary purpose of the Roster Test treatment was to test the household roster test question separately since changes in the amount and types of people included in the household could impact the results of person-level topics. Therefore, the analyses for Test Version 2 of the Health Insurance Coverage, Labor Force, and Income questions could have been impacted by these changes. However, it was determined that the additional information gained from testing an additional version of the topics in the Roster Test treatment was worth the risk.[5]

Table 1. Questions by Treatment

Topic	Control Treatment	Test Treatment	Roster Test Treatment
Household Roster	Production	Production	Test Version
Solar Panels	Test Version	Test Version	Test Version
Electric Vehicles	Test Version 1	Test Version 2	Test Version 1
Sewer	Test Version	Test Version	Test Version
Educational Attainment	Production	Test Version	Production
Health Insurance Coverage	Production	Test Version 1	Test Version 2
Disability	Production	Test Version	Production
SNAP	Production	Test Version	Test Version†
Labor Force	Production	Test Version 1	Test Version 2
Income	Production	Test Version 1	Test Version 2

† The SNAP Test Version will be in both test treatments to align with Labor Force and Income that also have a reference period change to the previous calendar year.

[5] We examined differences in key household and person characteristics among the Control and Roster Test treatments to explore any indication of bias in the Health Insurance Coverage, Labor Force, and Income analyses. See Spiers et al. (2023) for more information.

2.2 Data Collection

The 2022 ACS Content Test occurred in parallel with data collection activities for the September 2022 ACS production panel. Data collection for production ACS data consists of two main phases: an approximately two-month self-response data collection phase and a one-month follow-up phase.

During the self-response phase, addresses in sample are asked to self-respond by internet or mail. The Census Bureau sends addresses in sample up to five mailings to encourage self-response. This operation is followed by a one-month Computer-Assisted Personal Interviewing (CAPI) operation, where Census Bureau field representatives attempt to complete a survey for a sub-sample of the remaining nonresponding addresses.

The following data collection protocols for the 2022 ACS Content Test remained the same as production ACS:

- Data were collected using the self-response modes of internet (in English and Spanish) and paper questionnaires for the first and second month of data collection.
- In the third month of data collection, a sub-sample of nonresponding addresses were selected for CAPI.
- During CAPI, Census Bureau field representatives conducted interviews in person and over the phone.
- Self-response via internet or paper was accepted throughout the three-month data collection period.

The following data collection protocols for the 2022 ACS Content Test differed from production ACS:

- There were no paper versions of the 2022 ACS Content Test questionnaires in Spanish.[6]
- If respondents called Telephone Questionnaire Assistance (TQA) and opted to complete the survey over the phone, the interviewers conducted the survey using the production ACS questionnaire.[7] Since the TQA interviews did not include test questions, they were excluded from the analysis of the 2022 ACS Content Test.
- The 2022 ACS Content Test did not include the Telephone Failed-Edit Follow-Up (FEFU) operation. In production, this operation follows up on households that provided

[6] In 2019, 412 Spanish questionnaires were mailed back out of all mailable cases. Based upon this rate, we projected that only 8 Spanish questionnaires would be mailed back in the 2022 Content Test, which would not be cost-effective.

[7] The interviewer did not know which treatment the caller was in and therefore administered the production questionnaire. In 2019, less than one percent (0.6%) of cases responded by TQA and had no other response in a different mode. Based upon this rate, we projected about 744 TQA-only responses would be excluded from the 2022 ACS Content Test analysis.

incomplete information on the form or reported more than five people on the roster of a paper questionnaire.[8]

- The 2022 ACS Content Test used a telephone reinterview component to measure response reliability or response bias (depending upon the ACS topic). This telephone reinterview operation is discussed in Section 2.3 below.

For detailed information about ACS data collection procedures, consult the ACS Design and PRCS Design and Methodology Report (U.S. Census Bureau, 2022b).

2.3 Content Follow-Up Operation

To measure response reliability or response bias, a Content Follow-Up (CFU) reinterview was attempted with every household with an original Content Test interview that met the CFU eligibility requirements. Among the requirements were that the household must be occupied, and the household must have a valid telephone number. See the CFU requirements document for the complete list of eligibility requirements (Spiers, 2021a).

2.3.1 Content Test Follow-Up Protocol

As in previous ACS Content Tests, a case was sent to the CFU operation no sooner than two weeks (14 calendar days) after the original interview and had to be completed within three weeks after being sent to the CFU. This timing attempted to balance two competing needs: (1) to minimize the possibility of real changes in answers due to a change in life circumstances between the two interviews; (2) to minimize the possibility of the respondent repeating their previous answer based on their recollection of the original interview response, rather than considering the most appropriate answer.

All CFU reinterviews were conducted by telephone. At the first contact with a household, interviewers asked to speak with the original respondent. If that person was not available, interviewers scheduled a callback at a time when the original respondent was expected to be available. If this respondent could not be reached at the time of the second contact, the interviewer requested to speak with any other eligible household member (a household member who is 15 years or older). CFU reinterviews for the Content Test were conducted in either English or Spanish.

The CFU data collection instrument included the questions being tested for the 2022 ACS Content Test and some production ACS questions for context. It also included questions on public assistance from the 2022 Current Population Survey Annual Social and Economic

[8] The information obtained from the FEFU improves accuracy in a production environment but confounds the evaluation of respondent behavior in the Content Test environment. For paper questionnaires, where the household size is six or more (up to 12), we only collected name, age, and sex of these additional persons, but not detailed information as we do in the FEFU operation for ACS production.

Supplement (CPS ASEC) to measure response bias in the income from the public assistance question.

The CFU collected an independent household roster by re-asking the Household Roster questions along with Relationship, Sex, Age, and Date of Birth. The remaining CFU questions were only asked of the original household roster members. Only the Control and Roster Test panels collected an independent household roster. The Test panel used the original household roster to ask housing and detailed person questions.[9]

2.3.2 Content Test Follow-Up for Educational Attainment

The CFU will be used to answer Research Question 5: *Are the measures of response reliability (Gross Difference Rate, Index of Inconsistency) for the Control and Test treatments different?* We collapsed the "No schooling completed", "Nursery school", and "Kindergarten" response categories in the Control treatment into a single response category, "Less than grade 1." We measured response reliability for each treatment by comparing the responses in the original interview to the responses in the CFU reinterview (conducted via telephone). The analysis was conducted across mode to ensure that we had sufficient sample in the individual response categories. We tested for differences between the Control and Test treatments in the individual response categories using a t-test at the 0.1 level of significance.

2.4 Analysis Metrics

The sample addresses for the Control and Test treatments were selected in a manner so that their response propensities and response distributions (on particular characteristics) would be the same. Similar distributions allow us to conclude that any difference in the metrics used to analyze the educational attainment question is attributable to differences in the wording and format. We tested these unit-level assumptions in both the original interview and the CFU interview. See Section 2.4.1 for details. The metrics that we used to evaluate changes to the educational attainment question are presented in Section 2.4.2.

For the 2022 ACS Content Test, typical production ACS edits were not made because the primary concern of this test was how changes to existing questions and differences between versions of new questions affected the unaltered responses provided directly by respondents. For this reason, responses were not imputed either. A few edits were applied to the non-topic data, such as calculating a person's age based on his or her date of birth, but such edits were minimal.[10]

[9] The Test panel did not need to collect an independent household roster. The independent roster was needed to calculate the response reliability metrics for the Household Roster topic, which only used data from the Control and Roster Test treatments.

[10] This only refers to edits made to the data sets before analysis. During the analysis phase, additional edits, such as collapsing categories, were made based on the needs of the individual question.

All estimates from the ACS Content Test were weighted. The final Content Test weights account for the initial probability of selection (the base weight) and CAPI sub-sampling. The weights used in the CFU analysis also included an adjustment for CFU non-response.[11]

Comparisons between the Control and Test versions of educational attainment were conducted using a two-tailed t-test at the α=0.1 level of significance. The Content Test sample size was chosen to provide enough statistical power (0.80) to detect a difference in the gross difference rates (measuring differences in adds and deletes from the household roster) of at least two percentage points between the Control and Roster Test groups for the Household Roster question.[12] In statistical tests involving multiple comparisons, we controlled for the overall Type I error rate by adjusting the resulting p-values using the Hochberg method (Hochberg, 1988).[13]

We estimated the variances of the estimates using the Successive Differences Replication (SDR) method with replicate weights, the standard method used in the ACS (see U.S. Census Bureau, 2022b, Chapter 12). We calculated the variance for each rate and difference using the formula below. The standard error of an estimate (X_0) is the square root of the variance:

$$Var(X_0) = \frac{4}{80}\sum_{r=1}^{80}(X_r - X_0)^2$$

where:

X_0 = the estimate calculated using the full sample,
X_r = the estimate calculated for replicate r

2.4.1 Unit-Level Analysis

The unit response rate is important, as it provides an indication of the quality of the survey data. As part of our analysis, we examined unit-level (i.e., address-level) responses for the Control and Test treatments in the original interviews and CFU reinterviews. These results are provided in a separate report (Spiers et al., 2023).[14]

[11] The Content Test weight creation process does not include all the steps followed in the ACS, including the noninterview adjustment for the original interview and calibration to housing unit and population controls (see U.S. Census, 2022b, Chapter 11). For more information on the 2022 Content Test weighting procedure, see Risley and Oliver (2022) and Keathley (2022).

[12] See Section 2.4.2.4 for the definition of Gross Difference Rate.

[13] Use the MULTTEST Procedure in SAS®.

[14] As part of the 2022 ACS Content Test, we analyzed respondent burden. The results of this analysis are contained in Virgile et al. (2023).

2.4.2 Topic-Level Analysis

To evaluate the changes to educational attainment, we calculated a variety of metrics, presented in Sections 2.4.2.1 through 2.4.2.4.

2.4.2.1 Benchmarks

To roughly assess how well the educational attainment questions in the 2022 ACS Content Test estimated the percent of adults with a reported educational attainment less than grade 1, we compared the estimate for each treatment to a similar estimate from the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), an external reliable source (i.e., a benchmark). Our comparison was a nominal comparison only.

The benchmark measure came from the CPS ASEC, Table 1 of the CPS Educational Attainment tables: "Educational Attainment of the Population 18 Years and Over, by Age, Sex, Race, and Hispanic Origin."

2.4.2.2 Item Missing Data Rates

To measure nonresponse to the educational attainment question, we calculated its item missing data rate, the proportion of eligible persons (3 years of age or older) for which a required response is missing. A high item missing data rate can be indicative of a question that lacks clarity, is sensitive, or is simply too difficult to answer.

We compared item missing data rates by age group within mode for the Control and Test treatments via two-tailed t-tests.

2.4.2.3 Response Distributions

To assess how changes to educational attainment question affected the distributions for persons with educational attainment less than grade 1, we compared the percentage of persons with educational attainment less than grade 1 for the Control and Test treatments by age group within mode via two-tailed t-tests.

To assess changes to educational attainment question affected the distributions for the educational attainment levels shown below, we compared the proportions of the Control and Test treatment for these levels collectively by age within mode, using a Rao-Scott chi-square test that checks for a significant difference between two sample distributions (Rao & Scott, 1987).

- Less than grade 9
- Grades 9 through 12 (no diploma)
- High school graduate (regular high school diploma, GED or alternative credential)
- Some college credit, no degree
- Associate's degree
- Bachelor's degree or higher

If the chi-square test indicated a significant difference between the Control and Test distributions, we tested for significant differences in the individual category proportions using two-tailed t-tests.

2.4.2.4 Response Reliability

Survey responses are subject to error. Response error occurs for a variety of reasons, such as flaws in the survey design, misunderstanding of the questions, misreporting by respondents, and interviewer effects. For the 2022 ACS Content Test, response error was measured through response reliability or response bias, not both. This was done to reduce respondent burden and breakoffs during the CFU operation. For educational attainment, we measured response error using the Gross Difference Rate and Index of Inconsistency.

A survey question has good response reliability if respondents tend to answer the question consistently. For the 2022 ACS Content Test, we measured response reliability for a given question by comparing the responses to this question in the original interview to the responses to this same question in the CFU reinterview.

Re-asking the same question of the same respondent allows us to measure simple response variance, using the following measures:

- Gross difference rate (GDR)
- Index of inconsistency (IOI)
- L-fold index of inconsistency (IOI_L)

The first two measures, GDR and IOI, were calculated for individual response categories. The L-fold index of inconsistency was calculated for questions that had three or more mutually exclusive response categories, as a measure of overall reliability for the question.

In Table 2, “Yes” indicates that the unit is in the category of interest, according to the response from either the original interview or the CFU reinterview. “No” indicates that the unit is not reported to be in the category.

Table 2. Original Interview and CFU Reinterview Counts for Calculating GDR, IOI, and NDR

		Content Test original interview		reinterview totals
		Yes	No	
CFU reinterview	Yes	a	b	a + b
	No	c	d	c + d
original interview totals		a + c	b + d	n

Here, a, b, c, d, and n are counts, defined as follows:

a = units in category for both interview and reinterview
b = units not in category for original interview, but in category for reinterview
c = units in category for original interview, but not in category for reinterview
d = units in category for neither interview nor reinterview
n = total units in the universe = a + b + c + d

These counts were weighted to make them more representative of the population.

We calculated the GDR for this response category as:

$$GDR = \left(\frac{b+c}{n}\right) \times 100$$

To define the IOI, we must first discuss the variance of a category proportion estimate. If we are interested in the true proportion of a total population that is in a certain category, we can use the proportion of a survey sample in that category as an estimate. Under certain reasonable assumptions, it can be shown that the total variance of this proportion estimate is the sum of two components, sampling variance (SV) and simple response variance (SRV). It can also be shown that an unbiased estimate of SRV is half of the GDR for the category.

The SV is the part of total variance resulting from the differences between all the possible samples of size n one might have selected. SRV is the part of total variance resulting from the aggregation of response error across all sample units. If the responses for all sample units were perfectly consistent, then SRV would be zero, and the total variance would be due entirely to SV. As the name suggests, the IOI is a measure of how much of total variance is due to inconsistency in responses, as measured by SRV. A preliminary definition of the IOI is:

$$IOI = \left(\frac{SRV}{SRV + SV}\right) \times 100$$

We can estimate SRV using the GDR, but also need to estimate the denominator (i.e., total variance) in this expression. Based on previous studies, the estimate we use for total variance is:

$$SRV + SV = \frac{p_1 q_2 + p_2 q_1}{2}$$

where:

$$p_1 = \frac{a+c}{n} = \text{original interview proportion in category}$$

$$q_1 = 1 - p_1 = \frac{b+d}{n} = \text{original interview proportion not in category}$$

$$p_2 = \frac{a+b}{n} = \text{CFU proportion in category}$$

$$q_2 = 1 - p_2 = \frac{c+d}{n} = \text{CFU proportion not in category}$$

In comparing relative reliability (or response error) between treatments, if the response categories are essentially the same, then we looked at the differences in the GDR and IOI for each response category. We tested the significance of these differences, using two-tailed t-tests.

If the response categories did not match up exactly between the compared treatments, we either collapsed response categories to form equivalent categories for comparison, or we conducted comparisons for the response categories where it made sense.

So far, we have only discussed response reliability with respect to single response categories. If a question has three or more response categories (or "comparison categories" in cases where it is necessary to collapse some response categories for comparison), we also measured the overall response reliability of a question using the L-fold index of inconsistency, IOI_L. We looked at the difference in IOI_L between treatments and tested for significance as with the single category measures.

Suppose a question has L response categories. Let X_{ij} be the weighted count of sample units (households or persons) for which we have CFU responses in category *i* and original interview responses in category *j*. Here, both *i* and *j* range from 1 to L. Table 3 shows a cross-tabulation of the original interview and CFU results for a generic analysis topic. Note that if L = 2, then Table 3 is equivalent to Table 2.

Table 3. Cross-Tab of Original Interview and CFU Results: Questions with Response Categories

		Original Interview categories						
		1	2	…	j	…	L	CFU totals
CFU categories	1	X_{11}	X_{12}	…	X_{1j}	…	X_{1L}	X_{1+}
	2	X_{21}	X_{22}	…	X_{2j}	…	X_{2L}	X_{2+}
	…	…	…	…	…	…	…	…
	i	X_{i1}	X_{i2}	…	X_{ij}	…	…	X_{i+}
	…	…	…	…	…	…	…	…
	L	X_{L1}	X_{L2}	…	X_{Lj}	…	X_{LL}	X_{L+}
Original interview totals		X_{+1}	X_{+2}	…	X_{+j}	…	X_{+L}	$T = \sum_{i=1}^{L} \sum_{j=1}^{L} X_{ij}$

Now define the following proportions:

$$p_{ij} = \frac{X_{ij}}{T}$$

$$p_{+j} = \frac{X_{+j}}{T}$$

$$p_{i+} = \frac{X_{i+}}{T}$$

The IOI_L is calculated as

$$\mathrm{IOI_L} = \frac{1 - \sum_{i=1}^{L} p_{ii}}{1 - \sum_{i=1}^{L} (p_{i+} p_{+i})} \times 100$$

It can be shown that the IOI_L is a weighted sum of the L category IOI values (Biemer, 2011), but this formula is easier for calculation.

The IOI metrics can be biased if the parallel measures assumption is violated—i.e., the errors in the original interview and the CFU reinterview are positively or negatively correlated (Biemer, 2011). We checked this assumption by testing if the net difference rate (NDR) is significantly different from zero. The NDR is the difference between the original interview proportion of positive responses (“Yes” or in the category of interest) and the CFU proportion of positive responses. The NDR is calculated as follows:

$$NDR = (p_1 - p_2) \times 100 = \left(\frac{c - b}{n}\right) \times 100$$

If the NDR is significantly positive or negative, the assumption of "parallel measures" necessary for the SRV and IOI to be valid is not satisfied (Biemer, 2011). In these situations, we will use the following adjustment of the IOI, developed by Flanagan (2001):

$$\mathrm{IOI}_{\mathrm{adjusted}} = \frac{\frac{n^2(b+c) - n(c-b)^2}{n-1}}{(a+c)(c+d) + (a+b)(b+d)} \times 100$$

3 DECISION CRITERIA

Before field testing Educational Attainment, a team of subject matter experts identified and prioritized which of the research questions presented in Section 1.3.4, would determine which version of the educational attainment question be recommended for inclusion in the ACS. The decision criteria for Educational Attainment are presented in Table 4.

Table 4. Decision Criteria for Educational Attainment

Priority	Research Question	Decision Criteria
1	3	The Test version has a lower percentage of persons with educational attainment less than grade 1.
2	4	The distribution of educational attainment across response categories for the Control and Test versions are different.
3	1	The estimate of the percentage of the population with educational attainment less than grade 1 in the Test version is nominally close to the estimate in the CPS.
4	2	The missing data rate for the Test version is the same or lower than the missing data rate for the Control version.
5	5	The measures of response variance (GDR, IOI) for the Test version is the same or lower than that of the Control version in the less than grade 1 response category.

4 ASSUMPTIONS AND LIMITATIONS

4.1 Assumptions

- The sample addresses for the Control and Test treatments were selected in a manner so that their response propensities and response distributions would be the same. This assumption of homogeneity allows us to conclude that any difference between treatments is attributable to differences in wording and format. See Section 5 for more details.

- There was no difference between treatments in mail delivery timing or subsequent response time. The treatments had the same sample size and used the same postal sort and mailout procedures. Previous research indicated that postal procedures alone could cause a difference in response rates at a given point in time between experimental treatments of different sizes, with response for the smaller treatments lagging (Heimel, 2016).

- We assume that the frequency of real changes in answers due to a change in life circumstances between the original interview and CFU reinterview were similar between treatments.

4.2 Limitations

- GQs were not included in the sample for the 2022 ACS Content Test. The results of the Content Test may not extend to GQ populations.

- Housing units from Alaska, Hawaii, and Puerto Rico were not included in the sample for the 2022 ACS Content Test. The results of the Content Test may not extend to the housing unit population in these areas.

- The paper questionnaire was only available in English and was not available in Spanish like in production. The Content Test results related to the English paper questionnaire may not extend to Spanish paper questionnaire.

- For paper questionnaires, where the household size is six or more (up to 12), we only collected name, age, and sex of these additional persons. Detailed information for these persons in ACS production are collected in the FEFU operation. We did not include the FEFU operation because the information collected from it improves accuracy and could confound respondent behavior in the Content Test environment.

- We did not have response data for some partial internet responses (179 cases) due to a server issue. These cases were excluded from the analyses.

- TQA responses were excluded from the analysis of the 2022 ACS Content Test response data because survey responses completed via the TQA operation were only conducted using the ACS production data collection instrument.

- CAPI interviewers were assigned 2022 ACS Content Test cases as well as regular production cases. The potential risk of this approach is the introduction of a cross-contamination or carry-over effect among Control and test treatments and production due to the same interviewer administering multiple versions of the same question item (despite their training to read questions verbatim).

- Due to budget constraints, the CAPI workload could not exceed 28,000 housing units. This workload was less than what was subsampled originally because of over-sampled addresses in low response areas. Limiting the CAPI workload caused an increase in the variances for the analysis metrics used.

- The CFU reinterviews were conducted by phone only, whereas the original interviews were completed online, by mail, by phone in CAPI, and in person in CAPI. Hence, some of the differences observed between the original interviews and the CFU interviews may be the result of mode effect.

- Not all households who provided a response in the original interview were eligible for the CFU reinterview (see Section 2.3 for more information). As a result, 2.5 percent (standard error 0.2) of households from the original Control interviews, 2.5 percent (standard error 0.2) of households from the original Test interviews, and 3.0 percent (standard error 0.2) of households from the original Roster Test interviews were not eligible for the CFU reinterview. These rates were not significantly different between treatments (chi-square p-value 0.11).

- We reinterviewed the same person who responded in the original interview when possible but accepted interviewing a different person from the same household after two unsuccessful attempts at reaching the original person. Therefore, differences in results between the original interview and CFU reinterview for these cases could partly be from different people answering the questions. We interviewed a different household member in CFU for 7.3 percent (standard error 0.4) of CFU Control cases, 9.4 percent (standard error 0.5) of CFU Test cases, and 8.5 percent (standard error 0.5) of CFU Roster Test cases. These rates were significantly different between treatments (chi-square p-value 0.01) with the rate of CFU Test cases (t-test p-value <0.01) and CFU Roster Test cases (t-test p-value 0.04) being significantly higher than the rate of CFU Control cases.

- We examined potential differences between CFU respondents and nonrespondents within some socioeconomic and demographic characteristics because there were differences in the 2016 CFU reinterview (Spiers, 2021b). For all treatments combined, there were significant differences between CFU respondents and nonrespondents for *household size, tenure, age, race, Hispanic origin, language of original interview response,* and *high and low response areas*. These differences are similar to the ones found in the 2016 CFU (Spiers, 2021b).

- The 2022 ACS Content Test did not include the production weighting adjustments for unit nonresponse or population controls which are designed to minimize nonresponse and under-coverage bias. As a result, any estimates derived from the Content Test data did not provide the same level of inference as the production ACS and cannot be compared to production estimates.

5 RESULTS

This section of the report presents the results of various metrics used to evaluate Educational Attainment. The comparisons presented assume homogeneity of the response distributions for the three treatments, prior to the field test. We tested this assumption via unit-level (i.e., address level) analyses. The results are presented in Spiers et al.(2023).

In general, the overall unit response rates were not significantly different between treatments, nor were the response rate portions by mode. Additionally, when examining demographic and socioeconomic distributions, none of the response distributions were significantly different between treatments.

When looking at distributions among self-responses and CAPI responses, only the distribution for race in the “other race only” category among CAPI responses for the Control and Test treatments was significantly different. However, this is unlikely to have an impact on the educational attainment question.

There is no evidence of underlying CFU response rate issues that would negatively affect topic-level response error analyses. However, when examining demographic and socioeconomic distributions, the distributions for tenure were found to be different between the Control and Test treatments for those that responded during CFU and responded by self-response in the original interview. This difference was due to the Test treatment having a higher percentage of households reporting they own their household free and clear than the Control treatment.

5.1 Benchmark Results for Educational Attainment

RQ. 1. *How do the estimates of the percentage of the adult population with educational attainment “Less than grade 1” in the Control and Test treatments compare with the estimate in the Current Population Survey Annual Social and Economic Supplement (CPS ASEC)?*

The 2022 ACS Content Test results in Table 5 show that the estimate for the percentage of adults (18 or older) who report having an educational attainment "Less than grade 1” is 1.5 percent (standard error (SE)=0.1) for the Control treatment and 0.6 percent (SE=0.1) for the

Test treatment. The Test treatment percentage is significantly lower than that of the Control treatment.

Our benchmark data, from the 2022 educational attainment table package for the CPS ASEC, specifically, Table 1 from the table package, shows that 0.3 percent (SE=0.03) of adults (18 or older) have an educational attainment "Less than grade 1."[15] The Test treatment percentage is nominally closer in value to this benchmark measure than the Control treatment percentage. Based on the decision criteria, these results provide support for adopting the test version of the educational attainment question since the results from the Test version are closer to our benchmark.

Table 5. Percentage of Adults with Educational Attainment Less than Grade 1

Mode	Control	Test	Difference	P-value
Internet, Paper, CAPI	1.5(0.1)	0.6(0.1)	-0.9(0.1)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23- ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

5.2 Item Missing Data Rate Results for Educational Attainment

RQ. 2. *Is the missing data rate different for the Test treatment than for the Control treatment?*

Table 6 provides the results of the missing data rates for the Control and Test versions of the educational attainment question for the age groups within mode shown. The across mode results combine the results for the internet, paper, and CAPI modes.

The overall item missing data rate for all eligible persons (age≥3) across mode is significantly higher for the Test treatment compared to the Control treatment but the difference is relatively small (0.9 percentage points (SE=0.5)). The item missing data rates for the Test treatment are significantly higher than the Control treatment for all age group and mode categories except the following, where there is no significant difference: 3 to 17 (across mode), 18 or older (self-response mode), and 25 to 34 (CAPI mode).

These results were not expected and are not in line with the decision criteria for adopting the Test version of the educational attainment question. That said, analyses of breakoff rates (Table 15) and use of help screens (Table 14) between the control and test version of the educational attainment question show that respondent burden was not greater on the Test version compared with the Control version. We do not have reason to anticipate increased non-

[15] This corresponds specifically to the 'None' category in Table 1. Standard errors are not available in the table package but were calculated using CPS ASEC replicate weights.

response if the Test version of the educational attainment question is adopted. Moreover, this particular result was not among the highest priorities in the decision criteria.

Table 6. Item Missing Data Rate for Educational Attainment

Mode	Age Group	Control	Test	Difference	P-value
Across mode	age≥3	10.6(0.3)	11.6(0.3)	0.9(0.5)	0.06*
Across mode	3≤age≤17	16.9(0.8)	18.1(0.9)	1.3(1.2)	0.32
Across mode	age≥18	9.2(0.3)	10.1(0.3)	0.9(0.5)	0.05*
Self-Response	age≥18	10.0(0.4)	10.5(0.3)	0.5(0.5)	0.33
CAPI	age≥18	6.5(0.6)	8.7(0.8)	2.3(1.0)	0.03*
Across mode	25≤age≤34	10.4(0.7)	13.0(0.8)	2.6(1.0)	0.01*
Self-Response	25≤age≤34	11.2(0.8)	14.5(0.9)	3.3(1.1)	<0.01*
CAPI	25≤age≤34	7.9(1.4)	8.1(1.7)	0.2(2.2)	0.92
Across Mode (FB)$^{+}$	age≥18	2.6(0.4)	3.5(0.4)	0.9(0.5)	0.07*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. (FB)$^{+}$ means "foreign born." An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

5.3 Response Distribution Results for Educational Attainment

RQ. 3. *Is the percentage of persons with educational attainment "Less than grade 1" for the Control and Test treatments different?*

Table 7 provides the results of the response distribution for educational attainment "Less than grade 1" for the age groups within mode shown. The across mode results combine the results for the internet, paper, and CAPI modes.

The percentage of all eligible persons (age≥3) across mode with educational attainment "Less than grade 1" is significantly lower for the Test treatment compared to the Control treatment. The Test treatment has significantly smaller percentages of persons with educational attainment "Less than grade 1" for all categories except the following, where there is no significant difference: 18 or older (CAPI mode) and 25 to 34 (CAPI mode). These results provide strong support for adopting the Test version of the educational attainment question. Based on the decision criteria, these results are the most important in making our recommendation.

Table 7. Percentage of Persons with Educational Attainment Less than Grade 1

Mode	Age Group	Control	Test	Difference	P-value
Across mode	age≥3	6.3(0.2)	4.7(0.2)	-1.6(0.3)	<0.01*
Across mode	3≤age≤17	29.2(1.0)	24.3(0.9)	-4.9(1.4)	<0.01*
Across mode	age≥18	1.5(0.1)	0.6(0.1)	-0.9(0.1)	<0.01*
Self-Response	age≥18	1.8(0.2)	0.5(0.1)	-1.3(0.2)	<0.01*
CAPI	age≥18	0.8(0.2)	1.2(0.3)	0.4(0.3)	0.15
Across mode	25≤age≤34	0.8(0.2)	0.3(0.1)	-0.5(0.2)	<0.01*
Self-Response	25≤age≤34	1.0(0.2)	0.2(0.1)	-0.8(0.3)	<0.01*
CAPI	25≤age≤34	0.2(0.1)	0.6(0.4)	0.4(0.4)	0.32
Across Mode (FB)$^{+}$	age≥18	5.5(0.5)	2.9(0.5)	-2.6(0.7)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. (FB)$^{+}$ means "foreign born." An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

RQ. 4. *Are these distributions of educational attainment for the Test and Control treatments different?*

- Less than grade 9
- Grades 9 through 12 (no diploma)
- High school graduate (regular high school diploma, GED or alternative credential)
- Some college credit, no degree
- Associate's degree
- Bachelor's degree or higher

Table 8 provides the results of a Rao-Scott Chi-Square comparison of the Control and Test treatment's response distributions by age within mode for the six-category educational attainment levels shown above. The distributions for the Control and Test treatment are significantly different for all categories except the 25 to 34 age group in the CAPI mode.

Table 8. Educational Attainment by Mode and Age Group: Response Distributions Compared

Mode	Age Group	Response Distribution	Rao-Scott Chi-Square P-value
Across Mode	age≥3	Education Level	<0.01*
Across Mode	3≤age≤17	Education Level	----
Across Mode	age≥18	Education Level	<0.01*
Self-Response	age≥18	Education Level	<0.01*
CAPI	age≥18	Education Level	<0.01*
Across Mode	25≤age≤34	Education Level	0.05*
Self-Response	25≤age≤34	Education Level	<0.01*
CAPI	25≤age≤34	Education Level	0.67
Across Mode (FB)$^{+}$	age≥18	Education Level	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: An asterisk (*) indicates a statistically significant result based on a Rao-Scott Chi-Square test of independence at the α=0.1 level of significance. The (----) symbol indicates that the statistic could not be computed because one of the education cell levels had a frequency of zero. (FB)$^{+}$ means "foreign born."

The educational categories that contribute to the significant results shown in Table 8 are provided in Tables 9-11. These tables, broken out by mode (across mode, self-response mode, and CAPI mode, respectively), show the percentage of persons in a given age group who attained a given educational level. The distributions in these tables only reflect respondents with valid responses and excludes missing data.

The across mode results in Table 9 indicate significant results in the following age categories: age≥3 (three levels), age≥18 (four levels), and age≥18 (FB) (one level). The p-values have been adjusted within age category to control the familywise type I error rate.

There are two shifts in the distribution between the Control and Test treatments beyond the "Less than grade 1" that are worth noting. First, Table 9 shows a 1.6 percentage point decline in the percentage of people 18 and over reporting they are a "high school graduate" in the Test treatment compared to the Control treatment. Second, there is a 2.1 percentage point increase in the percent of adults who report "some college, no diploma" in the Test treatment compared to the Control treatment. Both shifts are more pronounced in the CAPI response mode (see Table 11). These results were unexpected, but do not provide strong evidence against adopting the Test version of the educational attainment question. These shifts may be related to changes in question wording discussed in the conclusion to the report.

Table 9. Educational Attainment by Age Group: Across Mode

Age Group	Education Level	Control	Test	Difference	Adj. P-value
age≥3	Less grade 9	16.7(0.3)	15.9(0.4)	-0.8(0.5)	0.24
	Grades 9-12 no diploma	7.6(0.2)	8.3(0.2)	0.7(0.3)	0.06*
	High school graduate	21.2(0.4)	19.9(0.3)	-1.3(0.4)	0.02*
	Some college no diploma	16.4(0.3)	18.1(0.3)	1.8(0.4)	<0.01*
	Associate's degree	7.3(0.2)	7.3(0.2)	0.1(0.3)	0.86
	Bachelor's or higher	30.9(0.5)	30.4(0.4)	-0.5(0.5)	0.80
3≤age≤17	Less grade 9	78.3(0.7)	78.7(0.8)	0.4(1.1)	0.83
	Grades 9-12 no diploma	20.9(0.8)	20.6(0.8)	-0.2(1.1)	0.83
	High school graduate	0.5(0.1)	0.4(0.1)	-0.1(0.2)	0.83
	Some college no diploma	0.2(0.1)	0.2(0.1)	0.1(<.1)	0.83
	Associate's degree	<0.1(<0.1)	<0.1(<0.1)	<0.1(<0.1	0.83
	Bachelor's or higher	<0.1(<0.1)	<0.1(<0.1)	<0.1 (<0.1)	0.83
age≥18	Less grade 9	3.7(0.2)	2.9(0.2)	-0.9(0.3)	<0.01*
	Grades 9-12 no diploma	4.8(0.2)	5.8(0.2)	1.0(0.3)	<0.01*
	High school graduate	25.6(0.4)	24.0(0.3)	-1.6(0.5)	<0.01*
	Some college no diploma	19.8(0.4)	21.8(0.4)	2.1(0.5)	<0.01*
	Associate's degree	8.8(0.3)	8.9(0.3)	<0.1(0.3)	0.89
	Bachelor's or higher	37.3(0.5)	36.7(0.5)	-0.6(0.6)	0.58
25≤age≤34	Less grade 9	2.0(0.3)	1.1(0.3)	-0.9(0.4)	0.15
	Grades 9-12 no diploma	3.6(0.4)	4.1(0.4)	0.5(0.5)	0.52
	High school graduate	21.3(0.8)	19.5(0.9)	-1.8(1.2)	0.52
	Some college no diploma	19.2(0.9)	21.1(1.0)	1.9(1.3)	0.52
	Associate's degree	10.0(0.6)	9.4(0.6)	-0.6(0.8)	0.52
	Bachelor's or higher	43.9(1.1)	44.8(1.2)	0.8(1.3)	0.52
age≥18 $(FB)^+$	Less grade 9	13.4(0.8)	11.7(0.8)	-1.7(1.0)	0.43
	Grades 9-12 no diploma	7.2(0.5)	10.0(0.6)	2.8(0.9)	<0.01*
	High school graduate	22.5(0.9)	20.2(0.9)	-2.3(1.2)	0.31
	Some college no diploma	13.4(0.6)	14.6(0.8)	1.2(0.9)	0.61
	Associate's degree	6.4(0.5)	6.5(0.5)	0.1(0.7)	0.96
	Bachelor's or higher	37.1(1.1)	37.0(1.1)	-0.1(1.4)	0.96

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance. $(FB)^+$ means "foreign born."

The results in Table 10 indicate the source of the difference for the age≥18 and 25≤age≤34 response distributions in the self-response mode in Table 8. Three levels of educational attainment indicate significant results in the 18 and older age group. One level of educational attainment indicates significant results in the 25 to 34 age group.

Table 10. Educational Attainment by Age Group: Self-Response Mode

Age Group	Education Level	Control	Test	Difference	Adj. P-value
age≥18	Less grade 9	2.9(0.2)	1.7(0.1)	-1.2(0.2)	<0.01*
	Grades 9-12 no diploma	3.7(0.2)	4.9(0.2)	1.2(0.3)	<0.01*
	High school graduate	22.8(0.4)	22.0(0.4)	-0.8(0.5)	0.45
	Some college no diploma	19.3(0.4)	20.6(0.4)	1.4(0.5)	0.02*
	Associate's degree	8.9(0.3)	8.9(0.3)	<0.1(0.4)	0.95
	Bachelor's or higher	42.5(0.6)	41.8(0.5)	-0.7(0.6)	0.57
25≤age≤34	Less grade 9	1.5(0.3)	0.5(0.1)	-1.1(0.3)	<0.01*
	Grades 9-12 no diploma	2.8(0.4)	3.1(0.4)	0.4(0.6)	0.62
	High school graduate	17.4(0.9)	16.6(0.9)	-0.8(1.2)	0.62
	Some college no diploma	17.1(1.0)	19.1(1.1)	2.0(1.3)	0.60
	Associate's degree	10.3(0.7)	9.0(0.7)	-1.3(1.0)	0.62
	Bachelor's or higher	50.9(1.2)	51.6(1.3)	0.7(1.5)	0.62

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

The results in Tables 11 indicate the source of the difference for age≥18 response distributions in the CAPI mode in Table 8. Two educational attainment categories, "high school graduate" and "some college no diploma" indicate significant differences in the 18 or older age category.

Table 11. Educational Attainment by Age Group: CAPI Mode

Age Group	Education Level	Control	Test	Difference	Adj. P-value
age≥18	Less grade 9	6.6(0.5)	6.9(0.6)	0.3(0.8)	0.82
	Grades 9-12 no diploma	8.6(0.6)	8.8(0.7)	0.2(0.8)	0.82
	High school graduate	34.9(1.1)	30.6(0.9)	-4.3(1.4)	<0.01*
	Some college no diploma	21.5(0.8)	26.0(0.9)	4.5(1.2)	<0.01*
	Associate's degree	8.5(0.6)	8.7(0.6)	0.3(0.9)	0.82
	Bachelor's or higher	19.9(1.2)	18.9(1.0)	-1.0(1.5)	0.82
25≤age≤34	Less grade 9	3.5(0.8)	3.0(0.9)	-0.5(1.2)	0.67
	Grades 9-12 no diploma	6.0(1.1)	6.8(1.2)	0.8(1.5)	0.67
	High school graduate	32.8(2.3)	27.9(2.4)	-5.0(3.0)	0.57
	Some college no diploma	25.5(1.7)	26.8(2.4)	1.3(3.2)	0.67
	Associate's degree	8.9(1.4)	10.6(1.4)	1.6(2.0)	0.67
	Bachelor's or higher	23.2(2.1)	24.9(2.5)	1.7(3.1)	0.67

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

5.4 Response Reliability Results for Educational Attainment

RQ. 5. *Are the measures of response reliability (Gross Difference Rate, Index of Inconsistency) for the Control and Test treatments different?*

For the 2022 ACS Content Test, we measured response reliability for the educational attainment question by re-asking the same question in the CFU reinterview, a few weeks after the original interview was conducted.

We measured response reliability for educational attainment via the Gross Difference Rate (GDR) and Index of Inconsistency (IOI). The results are presented in Tables 12 and 13, respectively. The p-values have been adjusted across educational attainment levels to control the familywise type I error rate.

The GDR results in Table 12 indicate that the GDR rate for the Test treatment is significantly lower than that of the Control treatment in the "Less than grade 1" educational attainment level. This means that there was a higher degree of consistency between the answers on the educational attainment question in the original interview and the CFU reinterview in the Test treatment than the Control treatment for the "Less than grade 1" educational attainment level. The GDR for the Test treatment is however significantly higher than the Control treatment for the "Master's degree" and "Doctorate degree" levels.

Table 12. Educational Attainment Level: Gross Difference Rates

Educational Attainment Level	Control	Test	Difference	Adj. P-value
Less than grade 1	0.8(0.1)	0.4(0.1)	-0.5(0.2)	0.04*
Grade 1 through 11	2.7(0.3)	2.1(0.2)	-0.7(0.4)	0.46
12th grade, no diploma	2.1(0.2)	2.5(0.3)	0.4(0.4)	0.65
Regular high school diploma	11.6(0.5)	11.9(0.5)	0.3(0.7)	0.65
GED or alternative credential	2.4(0.3)	2.6(0.3)	0.2(0.4)	0.65
Some college, but less than 1 year	11.5(0.5)	13.5(0.6)	1.9(0.8)	0.19
One year of college, no degree	13.3(0.6)	13.9(0.5)	0.5(0.9)	0.65
Associate's degree	4.2(0.4)	4.0(0.3)	-0.2(0.5)	0.65
Bachelor's degree	3.2(0.3)	3.5(0.3)	0.3(0.4)	0.65
Master's degree	1.6(0.2)	3.2(0.3)	1.6(0.4)	<0.01*
Professional degree	1.6(0.2)	2.2(0.2)	0.6(0.3)	0.32
Doctorate degree	1.0(0.2)	1.8(0.2)	0.8(0.3)	0.03*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

The IOI results in Table 13 indicate that the IOI rates for the Test treatment are significantly higher than that of the Control treatment in three educational attainment levels: Master's degree, Professional degree, and Doctorate degree. The IOI is the proportion of the total variance of a proportion estimate that is due to simple response variance. If the estimate of the index is less than 20, the response variance is low. If the estimate of the index is between 20 and 50, the response variance is moderate. If the estimate of the index is greater than 50, the response variance is high.

These results on response reliability provide support for the adoption of the Test version of the educational attainment question. They show that there was a higher degree of consistency between the answers on the educational attainment question in the original interview and CFU reinterview in the Test version compared with the Control version for the "Less than grade 1" category.

Table 13. Educational Attainment Level: Index of Inconsistency Rates

Educational Attainment Level	Control	Test	Difference	Adj. P-value
Less than grade 1	77.1(7.2)	63.7(15.2)	-13.4(17.3)	0.73
Grade 1 through 11	28.4(2.6)	21.3(2.1)	-7.1(3.4)	0.33
12th grade, no diploma	80.1(3.8)	72.8(5.6)	-7.4(6.2)	0.73
Regular high school diploma	30.6(1.3)	31.7(1.3)	1.1(1.7)	0.73
GED or alternative credential	37.9(3.8)	40.5(2.9)	2.6(4.9)	0.73
Some college, but less than 1 year	75.7(2.3)	80.6(2.3)	4.9(3.5)	0.73
One year of college, no degree	67.5(2.2)	71.0(2.1)	3.5(3.6)	0.73
Associate's degree	25.1(2.1)	25.9(1.8)	0.9(2.5)	0.73
Bachelor's degree	8.8(0.8)	9.8(0.7)	1.0(1,1)	0.73
Master's degree	8.6(1.1)	14.8(1.1)	6.2(1.7)	<0.01*
Professional degree	37.4(5.2)	67.6(4.5)	30.2(6.5)	<0.01*
Doctorate degree	23.8(4.2)	52.1(4.4)	28.3(5.6)	<0.01*

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

5.5 Other Metrics

This section of the report presents select results of a separate study on respondent burden for the 2022 ACS Content Test (Virgile et al., 2023). The results presented below pertain to the educational attainment question. These results are unweighted. See Virgile et al. (2023) for details.

Table 14 presents the access rates for the help screens for educational attainment. There is no significant difference between the access rates for the Control and Test treatments. The relatively low access rates for both treatments indicate that only a small percent of respondents sought additional information to help them understand and complete the educational attainment question.

Table 14. Help Screen Access Rates for Educational Attainment

Control	Test	Difference	P-value
0.4(0.1)	0.4(0.1)	<0.1(0.1)	0.76

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Table 15 presents the breakoff rates for educational attainment. There is no significant difference between the breakoff rates for the Control and Test treatments. The relatively low breakoff rates for both treatments indicate that only a small percent of respondents broke away from the survey upon encountering the educational attainment question.

Table 15. Breakoff Rates for Educational Attainment

Control	Test	Difference	P-value
0.1(<0.1)	0.1(<0.1)	<0.1(<0.1)	0.71

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0062. Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

6 CONCLUSIONS AND RECOMMENDATIONS

The purpose of this content testing was to evaluate if combining the lowest response options of the educational attainment question reduces the number of respondents who select the "No schooling completed" category. This testing was driven by ongoing research at the Census Bureau that show a relatively high percentage of people selecting the response category "No schooling completed" in the self-response modes of the ACS when compared to benchmarks. Among respondents who select "No schooling completed", there are likely adults who have completed some schooling, but are misidentifying their level of attainment. This test was intended to test a rewording of the response categories to reduce the number of people improperly selecting "No schooling completed."

The recommendation of the Education and Social Stratification Branch and educational attainment subcommittee is to adopt the Test version of the educational attainment question.

The results presented in this report support recommending the Test version of the educational attainment question for implementation in the ACS. The proposed changes to the response categories of the educational attainment question lower the percentage of people who select the lowest response category. The percentage of the population that reported educational attainment of "Less than grade 1" was 1.5 percent on the Control version and 0.6 percent on the Test version, showing a significant 0.9 percentage point decline (Table 5). The estimate of 0.6 percent of adults with "Less than grade 1" on the Test version also aligns more closely with our benchmark estimate of 0.3 percent from the Current Population Survey compared with the Control version.

The distributions of educational attainment between the Test and Control treatments are significantly different by mode and age group. We expected to see changes in the "Less than grade 1" category but we also see notable shifts in the "high school graduate" and "some college, no diploma" category. The percentage of adults reporting educational attainment of "high school graduate" is 1.3 percentage points lower on the Test version than on the Control version, and 4.3 percentage points lower when looking specifically at CAPI response mode. In addition, we also see a significant 2.1 percentage point increase in the "some college, no diploma" category on the Test version compared with the Control treatment. This group could be among those who mistakenly select "No schooling completed" on the production version of the questionnaire so it may be reasonable to see increases in the "some college, no diploma"

attainment categories. Together, the shifts in the "high school graduate" and "some college, no diploma" categories are greater than the shift we see in the "Less than grade 1" category.

Beyond the adjustments to the response categories, it is worth noting that these distributional shifts may be related to the change in the question wording from "*What is the highest degree or **level** of school this person has COMPLETED?"* on the Control version to "*What is the highest **grade** of school or degree this person has COMPLETED?*" Cognitive testing showed that the terms "grade" and "level" were generally interpreted similarly in the context of the educational attainment question (RTI International 2022a). However, some respondents understood them differently, as one noted "They are different. Grade is Kindergarten through 12th grade, and level I feel like is college or any type of degree" (RTI International 2022a). Some of the distributional shifts between the Control treatment and Test treatment versions of the educational attainment question were unexpected, but do not outweigh the overall benefits of the Test version.

Based on the decision criteria, most of the analyses results provided evidence to support the implementation of the Test treatment version over the Control treatment. The results and recommendations in this report were shared with the Educational Attainment subcommittee, and the National Center for Education Statistics (NCES) supports making this change.
The recommendation of the Education and Social Stratification Branch is to move forward with the implementation of the Test version of the educational attainment question.

7 ACKNOWLEDGMENTS

The 2022 ACS Content Test would not have been possible without the participation and assistance of many individuals from the Census Bureau and other agencies. We sincerely appreciate their contributions.

- Census Bureau staff in the American Community Survey Office, Application Development and Services Division, Center for Behavioral Science Methods, Decennial Information Technology Division, Decennial Statistical Studies Division, Field Division, National Processing Center, Population Division, and Social, Economic, and Housing Statistics Division.

- Representatives from other agencies in the Federal statistical system serving on OMB's Interagency Working Group for the ACS and the topical subcommittees formed by the Working Group for each topic tested in the 2022 ACS Content Test.

- Staff in OMB's Statistical and Science Policy Office.

8 REFERENCES

Biemer, P. (2011). *Latent class analysis of survey error.* John Wiley & Sons, Inc.

Crissey, S., Bauman, K., & Peterson, A. (2007). *Evaluation report covering educational attainment*. U.S. Census Bureau. Retrieved November 16, 2021, from https://www.census.gov/content/dam/Census/library/working-papers/2007/acs/2007_Crissey_02.pdf

Flanagan, P.E. (2001). Measurement errors in survey response. (Unpublished doctoral dissertation). University of Maryland, Baltimore County.

Heimel, S. (2016). *Postal tracking research on the May 2015 ACS panel*. U.S. Census Bureau. 2016 American Community Survey Research and Evaluation Report Memorandum Series #ACS16-RER-01.

Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 75 (4), 800-802. Retrieved January 17, 2017, from https://doi.org/10.2307/2336325

Keathley, D. (2022). *Specifications for selecting and weighting the 2022 American Community Survey Content Test sample.* U.S. Census Bureau. DSSD 2022 American Community Survey Memorandum Series #ACS22-S-21.

Lee, J., & Zhou, M. (2015). Immigration hyper-selectivity and second-generation convergence. In *The Asian American achievement paradox* (pp. 21-49). Russel Sage Foundation.

Office of Management and Budget (2006). *Standards and guidelines for statistical surveys*. Retrieved February 24, 2022, from https://www.whitehouse.gov/wpcontent/uploads/2021/04/standards_stat_surveys.pdf

Rao, J., & Scott A. (1987). On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics*, 15 (1), 1987, 385–397. https://doi.org /10.1214/aos/1176350273

Risley, M., & Oliver, B. (2022). *High-level sampling and weighting requirements for the 2022 American Community Survey Content Test.* U.S. Census Bureau. DSSD 2022 American Community Survey Memorandum Series #ACS22-MP-01.

RTI International (2022a). *2022 ACS Content Test: Round 1 and Round 2 Cognitive Testing.* Retrieved July 19, 2022, from https://www.census.gov/content/dam/Census/library/working-papers/2022/acs/2022_Wilson_01.pdf

RTI International (2022b). *Cognitive Testing for the 2022 ACS Content Test: Round 3 Briefing* Retrieved August 29, 2022 from https://www.census.gov/library/working-papers/2022/acs/2022_Wilson_02.html

Spiers, S. (2021a). *Requirements for the Content Follow-Up Reinterview Survey in the 2022 American Community Survey Content Test*. U.S. Census Bureau. American Community Survey Memorandum Series #ACS22-MP-02.

Spiers, S. (2021b). Coverage and nonresponse bias in the 2016 American Community Survey Content Follow-Up reinterview. *Proceedings of the 2021 Joint Statistical Meetings, Survey Research Methods Section* (pp. 1664-1677). American Statistical Association.

Spiers, S., Baumgardner, S., & Mojica, A. (2023). *2022 American Community Survey Content Test evaluation report: Unit-level response*. Retrieved November 27, 2023, from 2022 American Community Survey Content Test Evaluation Report: Unit-Level Response (census.gov)

U.S. Census Bureau. (2022a). *U.S. Census Bureau statistical quality standards*. Retrieved September 16, 2022, from https://www2.census.gov/about/policies/quality/quality-standards.pdf

U.S. Census Bureau. (2022b). *American Community Survey and Puerto Rico Community Survey Design and Methodology, Version 3.0.* Retrieved December 15, 2022, from https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

Virgile, M., Weng, C. F., Mills, G., & Spiers, S. (2023). *2022 American Community Survey Content Test evaluation report: Respondent burden.* Retrieved November 27, 2023, from 2022 American Community Survey Content Test Evaluation Report: Respondent Burden (census.gov)