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

November 7, 2017

2017 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS17-RER-15

MEMORANDUM FOR	Victoria Velkoff Chief, American Community Survey Office
From:	David Waddington Chief, Social, Economic, and Housing Statistics Division (SEHSD)
Prepared by:	Anthony Martinez Social, Economic, and Housing Statistics Division (SEHSD)
Subject:	2016 American Community Survey Content Test Evaluation Report: Industry and Occupation

Attached is the final report for the 2016 American Community Survey (ACS) Content Test for Industry and Occupation. This report describes the results of the test for a revised version of the Industry and Occupation questions.

If you have any questions about this report, please contact Lynda Laughlin at 301-763-6314, Anthony Martinez at 301-763-3595, or Ana Montalvo at 301-763-5977.

Attachment

cc:
Agnes Kee (ACSO)
Jennifer Ortman (ACSO)
Brett Moran (CSRM)
Patrick Cantwell (DSSD)
Broderick Oliver (DSSD)
Elizabeth Poehler (DSSD)
Anthony Tersine (DSSD)
Jennifer Day (SEHSD)
Lynda Laughlin (SEHSD)
Nicole Scanniello (SEHSD)

Intentionally Blank

2016 American Community Survey Content Test Evaluation Report: Industry and Occupation

FINAL REPORT

Anthony Martínez
Ana J. Montalvo
Social, Economic, and Housing Statistics Division
Broderick Oliver
Decennial Statistical Studies Division

Intentionally Blank

TABLE OF CONTENTS

List of Tables

List of Figures

EXECUTIVE SUMMARY

Overview

From February to June of 2016, the U.S. Census Bureau conducted the 2016 American Community Survey (ACS) Content Test, a field test of new and revised content. The primary objective was to test whether changes to question wording, response categories, and definitions of underlying constructs improve the quality of data collected. Both new and revised versions of existing questions were tested to determine if they could provide data of sufficient quality compared to a control version as measured by a series of metrics including item missing data rates, response distributions, comparisons with benchmarks, and response error. The results of this test will be used to help determine the future ACS content and to assess the expected data quality of revised questions and new questions added to the ACS.

The 2016 ACS Content Test consisted of a nationally representative sample of 70,000 residential addresses in the United States, independent of the production ACS sample. The sample universe did not include group quarters, nor did it include housing units in Alaska, Hawaii, or Puerto Rico. The test was a split-panel experiment with one-half of the addresses assigned to the control treatment and the other half assigned to the test treatment. As in production ACS, the data collection consisted of three main data collection operations: 1) a six-week mailout period, during which the majority of self-response via internet and mailback were received; 2) a one-month Computer-Assisted Telephone Interview period for nonresponse follow-up; and 3) a one-month Computer-Assisted Personal Interview period for a sample of the remaining nonresponse. For housing units that completed the original Content Test interview, a Content Follow-Up telephone reinterview was conducted to measure response error.

Industry and Occupation

This report presents the test results of a new version of the questions on Industry and Occupation. A question about a worker's occupation has been asked on the census since 1820 and a question about the worker's industry has been asked since 1910. The current versions of these questions have been asked since 1960.

Ongoing research on the Industry and Occupation write-in responses has demonstrated that the current questions may be confusing. To improve occupational specificity, the Industry and Occupation questions were revised to include new and consistent examples and modified question wording. The number of characters permitted for responses to the Job Duties question was expanded from 60 to 100 characters. Additional lines were provided on the paper questionnaire; the internet version and computer-assisted interviewer modes had a larger box for the write-ins to accommodate this change.

Research Questions and Results

To evaluate the performance of the test version of the Industry and Occupation questions, we examined the following:

Item Missing Data Rates: We expected the item missing data rates to be lower in the test treatment for both Industry and Occupation. Overall, the test treatment had no effect. However, in the internet mode, the item missing data rate for Industry was significantly higher in the test treatment by 0.6 percentage points.

Codeable Rates: For Industry, the codeable rate was 0.5 percentage points lower for the test treatment in the mail mode. For Occupation, the codeable rate was significantly lower in the test treatment both overall and in the mail mode (0.4 and 0.8 percentage points lower, respectively). The overall codeable rate exceeds 98.0 percent for both Industry and Occupation in both treatments.

Interim Referral Rates: When a coding clerk is unable to code Industry, Occupation, or both, the case is referred to an expert coder (i.e., interim referral). The interim referral rates for the test treatment were significantly higher overall and for all modes except internet. In the internet mode, the difference between the interim referral rates for the test and control treatments was not significant.

Response Distributions: Overall, there was no significant difference between treatments in the response distributions for the Standard Occupational Classification (SOC) and the North American Industry Classification System (NAICS) major groups. However, for the NAICS major group, the response distributions differed in the internet mode and the CATI/CAPI combined mode.

Reliability: The response variance for the Military Industry was significantly lower for the test treatment by 27.8 percentage points, most likely attributable to moving the Armed Forces checkbox from the Employer Name to the Class of Worker question.[1] For Occupation, the response error was significantly lower in the test treatment for Personal Care and Service Occupations. For Industry, the response error was significantly lower for Other Public Services, except Public Administration.

Mean Character and Word Count, and Specificity of Write-In Responses: The increase in the number of characters allowed for responses to the Job Duties question in the test treatment resulted in a significant increase in the mean word and character count of responses overall and for all modes. The test treatment write-in responses for Job Duties, which are vital for coding, were more detailed in the test treatment than the control treatment.

Median Coding Time: We expected median coding time to assign a valid industry and occupation code would be less in the test treatment due to an expected increase in specificity in the write-in

[1] During the 2016 ACS Content Test, the Class of Worker and the Industry and Occupation questions were tested concurrently. Results are presented in separate reports.

responses. However, there was no significant difference in the median coding time between the test and control treatments.

Industry and Class of Worker Consistency Check: The consistency rate for *Private households* was significantly lower overall and in the internet mode for the test treatment. There was no significant differences in the remaining consistency rates between the control and test treatments.

Conclusion

The 2016 ACS Content Test proposed changes to the Industry and Occupation questions yielded a combination of results. Overall, the test treatment had a significantly lower codeable rate for occupation and a significantly higher interim referral rate. There were no significant differences in the overall results of the item missing data rates, response distributions, and median coding times between the treatments. One of the main reasons for testing a new version of the Industry and Occupation questions was to improve occupational specificity in the write-in responses. Results from the reliability metrics show the test treatment improved the clarity of both questions and increased the specificity of responses about occupation. The increased specificity of the write-in responses may make it possible to publish more refined industry and occupation detailed codes in the future. These enhancements are expected to improve the overall quality of the industry and occupation data and related data products available from the U.S. Census Bureau.

The recommendation of the Industry and Occupation Statistics Branch is to adopt the test version of the Industry and Occupation questions. The results of the 2016 ACS Content Test, most notably the increased specificity of the write-ins responses, may make it possible to produce more accurate and detailed codes for industry and occupation categories than is currently available using the current coding system for industry and occupation.

Intentionally Blank

1 BACKGROUND

From February to June of 2016, the Census Bureau conducted the 2016 American Community Survey (ACS) Content Test, a field test of new and revised content. The primary objective was to test whether changes to question wording, response categories, and definitions of underlying constructs improve the quality of data collected. Both revised versions of existing questions and new questions were tested to determine if they could provide data of sufficient quality compared to a control version as measured by a series of metrics including item missing data rates, response distributions, comparisons with benchmarks, and response error. The results of this test will be used to help determine the future ACS content and to assess the expected data quality of revised questions and new questions added to the ACS.

The 2016 ACS Content Test included the following topics:

- Relationship
- Race and Hispanic Origin
- Telephone Service
- Computer and Internet Use
- Health Insurance Coverage
- Health Insurance Premium and Subsidy (new questions)
- Journey to Work: Commute Mode
- Journey to Work: Time of Departure for Work
- Number of Weeks Worked
- Class of Worker
- Industry and Occupation
- Retirement, Survivor, and Disability Income

This report discusses the topic of Industry and Occupation.

1.1 Justification for Inclusion of Industry and Occupation in the Content Test

A question on a worker's occupation has been asked on the decennial census since 1820. A question that asks for information about the worker's industry has been asked since 1910. These questions have changed as the characteristics of the labor force and data collection processes at the Census Bureau have changed. The current versions of the questions have been asked since the 1960 Census.

Occupational Specificity

The ACS is the primary Census Bureau-sponsored source of information on occupations and the only source with the geographic detail needed to support the data needs of federal agencies and other data users. Multiple federal agencies, businesses, and other data users have requested greater occupational detail in our code list and product tables. However, we currently cannot accommodate those requests as data disclosure avoidance procedures prevent us from publishing data on occupations under a certain threshold in order to protect the confidentiality of our survey

respondents.[2] The proposed format for the Industry and Occupation questions may increase coding specificity and may make it possible to publish data on occupations that had not previously been disclosed due to small counts.

In addition, we anticipate the level of detail for the 2018 Standard Occupational Classification (SOC) will increase for several fields.[3] Significant updates are expected in the fields of *Management*, *Business*, *Finance*, *Information Technology*, *Engineering*, *Social Science*, *Education*, *Media*, *Healthcare*, *Personal Care*, *Extraction*, and *Transportation* Occupations (U.S. Office of Management and Budget, 2016). Thus, obtaining more detailed information on certain occupations will help the ACS to adhere to the current and potential future SOC requirements

Increasing Clarity of Industry and Occupation Questions

The current questions provide examples to help respondents answer the questions. However, these examples do not function as intended. The examples, which are lengthy, can confuse and irritate respondents (Raglin, 2014). In addition, field representatives find the examples, which they are required to be read out loud, cumbersome. Some respondents mistake the examples for the full set of possible outcomes and leave the field blank because they think that none of the examples applies to them.

The 2006 ACS Content Test tested changes to the examples, but did not compare differences between coding and distributions for the treatments (Tegler, Downs, Kirk, & Ericson, 2007). These comparisons were conducted in the 2016 ACS Content Test. Through cognitive testing, we learned that simply omitting examples was not necessarily the best option. Stapleton & Steiger (2015) found that having examples was helpful. In the 2016 ACS Content Test, we tested new examples to standardize content across the entire series of questions.

1.2 Question Development

Initial versions of the new and revised questions were proposed by federal agencies participating in the U.S. Office of Management and Budget (OMB) Interagency Committee for the ACS. The initial proposals contained a justification for each change and described previous testing of the question wording, the expected impact of revisions to the time series and the single-year as well as five-year estimates, and the estimated net impact on respondent burden for the proposed revision.[4] For proposed new questions, the justification also described the need for the new data, whether federal law or regulation required the data for small areas or small population groups, if other data sources were currently available to provide the information (and why any alternate sources were insufficient), how policy needs or emerging data needs would be addressed through the new question, an explanation of why the data were needed with the geographic precision and

[2] For more information on ACS Data Disclosure Avoidance Procedures, see https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch13_2014.pdf.

[3] For more information on the Standard Occupational Classification Manual (SOC), see https://www.bls.gov/soc/.

[4] The ACS produces both single and five-year estimates annually. Single year estimates are produced for geographies with populations of 65,000 or more and five-year estimates are produced for all areas down to the block-group level, with no population restriction.

frequency provided by the ACS, and whether other testing or production surveys had evaluated the use of the proposed questions.

The Census Bureau and OMB, as well as the Interagency Council on Statistical Policy Subcommittee, reviewed these proposals for the ACS. OMB determined which proposals moved forward into cognitive testing. After OMB approval of the proposals, topical subcommittees were formed from the OMB Interagency Committee for the ACS, which included all interested federal agencies that use the data from the impacted questions. These subcommittees further refined the specific proposed wording that was cognitively tested.

The Census Bureau contracted with Westat to conduct three rounds of cognitive testing. The results of the first two rounds of cognitive testing informed decisions on specific revisions to the proposed content for the stateside Content Test (Stapleton & Steiger, 2015). In the first round, 208 cognitive interviews were conducted in English and Spanish and in two modes (self-administered on paper and interviewer-administered on paper). In the second round of testing, 120 cognitive interviews were conducted for one version of each of the tested questions, in English and Spanish, using the same modes as in the first round.

A third round of cognitive testing involved only the Puerto Rico Community Survey (PRCS) and Group Quarters (GQ) versions of the questionnaire (Steiger et al., 2015b). Cognitive interviews in Puerto Rico were conducted in Spanish; GQ cognitive interviews were conducted in English. The third round of cognitive testing was carried out to assess the revised versions of the questions in Spanish and identify any issues with questionnaire wording unique to Puerto Rico and GQ populations.[5] The proposed changes identified through cognitive testing for each question topic were reviewed by the Census Bureau, the corresponding topical subcommittee, and the Interagency Council on Statistical Policy Subcommittee for the ACS. The OMB then provided final overall approval of the proposed wording for field testing.[6]

For Industry and Occupation, two versions of the questions were tested on paper in the first round of cognitive testing. There were slight wording differences between the two versions of the Industry question. One included examples, while the other provided an instruction to "be as specific as possible" with no examples to address concerns that providing examples could bias response.

New industry examples were tested in the version that included examples. Respondents could interpret that none of the examples apply to them and subsequently leave the field entirely blank or write-in one of the examples provided even if it was not an accurate response. Most respondents thought the examples were helpful and for some questions, such as Job Duties, the examples seemed to elicit more detail. Steiger et al. (2014) also found that the wording of the Job Title question was problematic for those with no official industry title; those whose title is specific to their industry, but not really descriptive of what they do; and those whose title does not match their occupation. The recommendation going into the second round of cognitive

[5] Note that the field testing of the content was not conducted in Puerto Rico or in GQs. See the Methodology section for more information.

[6] A cohabitation question and domestic partnership question were included in cognitive testing but ultimately we decided not to move forward with field testing these questions.

testing was for Kind of Business, Job Title, and Job Duties to use a combination of the question wording from the second version with examples from the first version of the questions (Steiger et al., 2014).

For the Job Duties question in the mail mode, Westat recommended that we continue to offer three lines for write-in responses as done in the second version of Round 1. Providing three lines, instead of one, for open-ended responses seemed to prompt most respondents to put in more detail, which makes it easier for clerical coders to interpret and classify responses. This expanded the number of characters allowed in the write-in responses from 60 to 100. There would be no perceptible difference for those who were administered the test treatment in the Computer-Assisted Interview (CAI) modes.

In Round 2, Westat tested the three questions with examples, along with revisions to the wording in the Job Title question (eliminating the word "title") and the Job Duties question (asking respondents to "describe" their duties). Westat also recommended using two consistent examples between the Kind of Business, Job Title, and Job Duties questions, one of which should be for a blue-collar occupation based on feedback from the first round of cognitive testing (Steiger et al., 2015a). The current occupation examples, registered nurse and patient care, are too general and lead to vague or unspecific write-ins, especially for *Healthcare* and *Healthcare-related* occupations. This is a particular issue for the Job Duties question and the current example patient care. Numerous respondents who work in healthcare will write in patient care, which is a vague job duty and reduces our ability to code *Healthcare* occupations with more specificity. For this round of testing, for Job Title we selected the examples of a *Teaching* occupation (4th grade teacher) and a *Construction and Extraction* occupation (entry-level plumber), with their respective duties as examples for the Job Duties question.

In addition, the Class of Worker, Industry, and Occupation items were tested in the second round as a single, multi-part question to show that these were all a part of a series with one general instruction format. Most respondents interpreted the series as intended, providing answers that reflected their job situation (Steiger et al., 2015a).

Westat suggested adding a closing phrase to enhance interviewer-administered example lists and to emphasize that the lists are not exhaustive, but were simply provided to give a better sense of how the questions fit together. For consistency across items, Westat recommended adding across all modes the phrase "or something else" to all the Industry and Occupation items to help respondents understand that non-listed responses were welcome, even for questions that were not deemed to be difficult for respondents (Steiger et al., 2015a). Heading into the third round of cognitive testing, it was decided that the phrases "or another kind of business," "or some other occupation," and "or other duties" text would only be used for the interviewer-administered modes as a closing statement due to the awkwardness of reading examples without a closing statement. In the mail mode, cognitive testing without these phrases appeared to work well. Adding the phrases to the examples would be inconsistent with example listings on other content on the mail questionnaire. Also, past experience has shown that many respondents would literally write one of these phrases in these write-in fields if this text were listed at the end of the examples in the self-administered (mail or internet) modes (Steiger et al., 2015a).

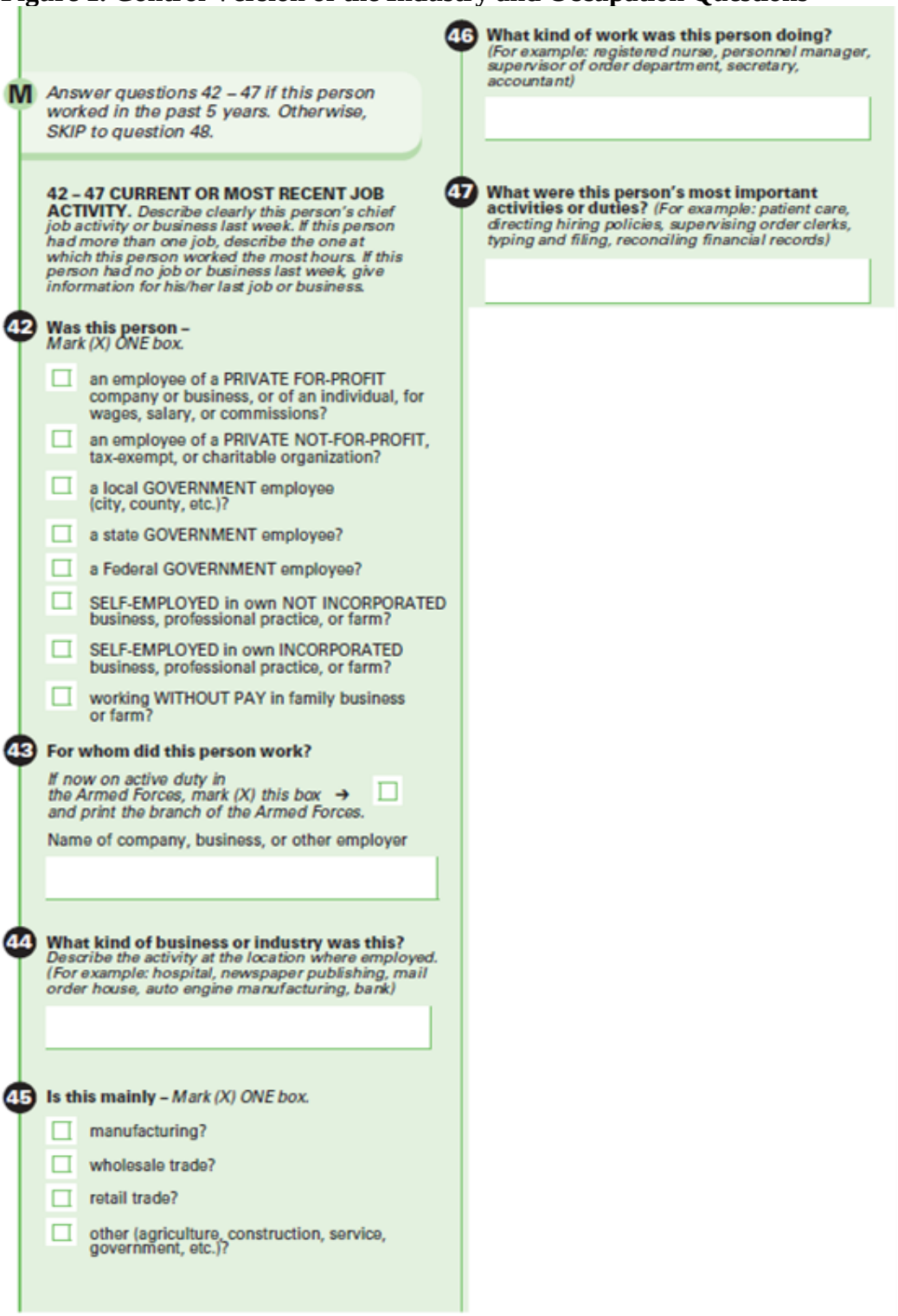
Westat also recommended that respondents be asked to "describe" their main occupation or title in the Job Title question, rather than asking what the occupation or title was. They also suggested removing the instruction to "Be as specific as possible." This instruction had two interpretations, on the one hand to provide as much detail as possible and on the other to be as precise and concise as possible. Neither interpretation seemed to be helpful in the context of providing the information requested, therefore the phrase was removed from the question (Steiger et al., 2015a).

1.3 Question Content

The control and test versions of the questions as they appeared on the mail questionnaire for the 2016 ACS Content Test are shown in Figures 1 and 2, respectively. Automated versions of the questionnaire had the same content formatted appropriately for each mode. For each treatment, we note the placement of the Class of Worker question in the figure.

For the question about Job Duties, the test version (Figure 2, Question 41f) provided more lines on the mail questionnaire than the control (Figure 1, Question 47). The internet version for the test treatment had a larger box for the write-ins (see Appendix A). Although the Computer-Assisted Telephone Interview (CATI) and Computer-Assisted Personal Interview (CAPI) fields in the test version were able to capture a greater character length, the respondent was not visually cued to this (see Appendix B).

Figure 1. Control Version of the Industry and Occupation Questions[7]



M *Answer questions 42 – 47 if this person worked in the past 5 years. Otherwise, SKIP to question 48.*

42 – 47 CURRENT OR MOST RECENT JOB ACTIVITY. *Describe clearly this person's chief job activity or business last week. If this person had more than one job, describe the one at which this person worked the most hours. If this person had no job or business last week, give information for his/her last job or business.*

42 **Was this person –**
Mark (X) ONE box.

- ☐ 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 NOT INCORPORATED business, professional practice, or farm?
- ☐ SELF-EMPLOYED in own INCORPORATED business, professional practice, or farm?
- ☐ working WITHOUT PAY in family business or farm?

43 **For whom did this person work?**

If now on active duty in the Armed Forces, mark (X) this box → ☐ *and print the branch of the Armed Forces.*

Name of company, business, or other employer

44 **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)

45 **Is this mainly –** *Mark (X) ONE box.*

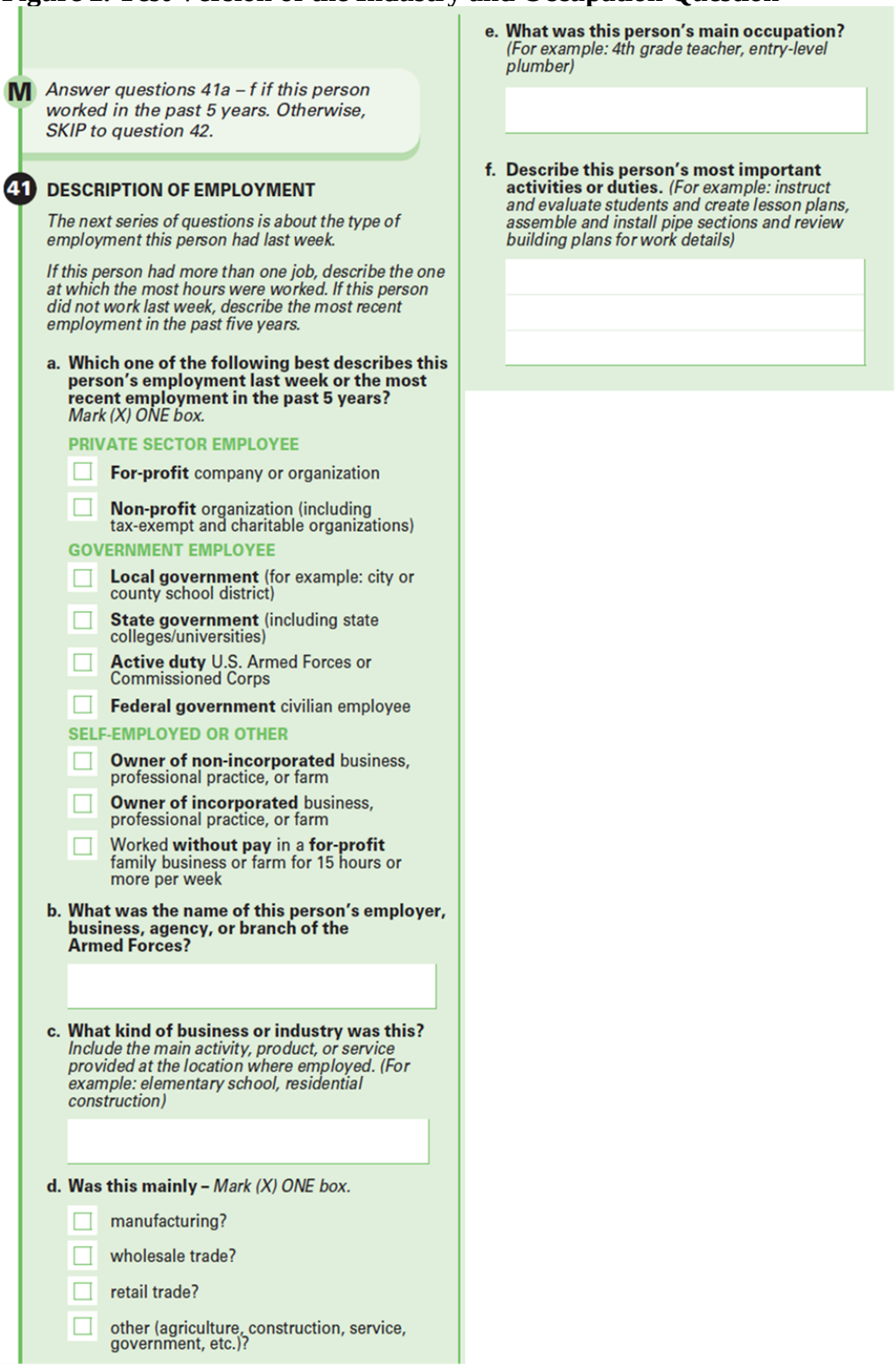
- ☐ manufacturing?
- ☐ wholesale trade?
- ☐ retail trade?
- ☐ other (agriculture, construction, service, government, etc.)?

46 **What kind of work was this person doing?**
(For example: registered nurse, personnel manager, supervisor of order department, secretary, accountant)

47 **What were this person's most important activities or duties?** *(For example: patient care, directing hiring policies, supervising order clerks, typing and filing, reconciling financial records)*

[7] Question 42 refers to Class of Worker. Both the Class of Worker and Industry and the Industry and Occupation questions were tested concurrently in the 2016 ACS Content Test. Results for the Class of Worker topic are published in a separate report.

Figure 2. Test Version of the Industry and Occupation Question[8]



M *Answer questions 41a – f if this person worked in the past 5 years. Otherwise, SKIP to question 42.*

41 **DESCRIPTION OF EMPLOYMENT**

The next series of questions is about the type of employment this person had last week.

If this person had more than one job, describe the one at which the most hours were worked. If this person did not work last week, describe the most recent employment in the past five years.

a. **Which one of the following best describes this person's employment last week or the most recent employment in the past 5 years?** *Mark (X) ONE box.*

PRIVATE SECTOR EMPLOYEE

- ☐ **For-profit** company or organization
- ☐ **Non-profit** organization (including tax-exempt and charitable organizations)

GOVERNMENT EMPLOYEE

- ☐ **Local government** (for example: city or county school district)
- ☐ **State government** (including state colleges/universities)
- ☐ **Active duty** U.S. Armed Forces or Commissioned Corps
- ☐ **Federal government** civilian employee

SELF-EMPLOYED OR OTHER

- ☐ **Owner of non-incorporated** business, professional practice, or farm
- ☐ **Owner of incorporated** business, professional practice, or farm
- ☐ Worked **without pay** in a **for-profit** family business or farm for 15 hours or more per week

b. **What was the name of this person's employer, business, agency, or branch of the Armed Forces?**

c. **What kind of business or industry was this?** *Include the main activity, product, or service provided at the location where employed. (For example: elementary school, residential construction)*

d. **Was this mainly –** *Mark (X) ONE box.*

- ☐ manufacturing?
- ☐ wholesale trade?
- ☐ retail trade?
- ☐ other (agriculture, construction, service, government, etc.)?

e. **What was this person's main occupation?** *(For example: 4th grade teacher, entry-level plumber)*

f. **Describe this person's most important activities or duties.** *(For example: instruct and evaluate students and create lesson plans, assemble and install pipe sections and review building plans for work details)*

[8] Question 41a refers to Class of Worker. Both the Class of Worker and Industry and the Industry and Occupation questions were tested concurrently in the 2016 ACS Content Test. Results for the Class of Worker topic are published in a separate report.

1.4 Research Questions

The following research questions were formulated to guide the analyses of the Industry and Occupation questions. The analyses assess how the test version of the questions performed compared to the control version in the following ways: how often the respondents answered the question, the consistency and accuracy of the responses, and how the responses affect the resulting estimates.

1. Are the control and test *item missing data rates* for **Industry** the same across data collection mode and within data collection mode?

2. Are the control and test *item missing data rates* for **Occupation** the same across data collection mode and within data collection mode?

3. Are the control and test *codeable data rates* for **Industry** the same across data collection mode and within data collection mode?

4. Are the control and test *codeable data rates* for **Occupation** the same across data collection mode and within data collection mode?

5. Are the control and test *interim referral rates* for **Industry/Occupation** the same across data collection mode and within data collection mode?

6. Are the control and test *distributions of eligible persons among the **NAICS** Industry sectors* (as indicated by the industry code) the same across data collection mode and within data collection mode?

7. Are the control and test *distributions of eligible persons among the **SOC** major groups* (as indicated by the occupation code) the same across data collection mode and within data collection mode?

8. Is *response reliability* for **Industry** better in test than control?

9. Is *response reliability* for **Occupation** better in test than control?

10. For each of the four **Industry and Occupation** write-in fields, is the *mean character count* for test greater than for control? (overall and by mode)

11. For each of the four **Industry and Occupation** write-in fields, is the *mean word count* for test greater than for control? (overall and by mode)

12. Do the changes to the **Industry and Occupation** questions result in more *specificity* in the four write-in responses across data collection mode and within data collection mode?

13. Is the *median coding time* for **Industry/Occupation** for test less than control?

14. For the test treatment, how consistent are **Class of Worker** responses with write-in responses about **Industry** (Employer Name and Kind of Business) and the final industry code compared to control? In particular, is the reporting of Active Duty in the Class of Worker test version consistent with the Industry write-in responses and industry code? (overall and by mode)

2 METHODOLOGY

2.1 Sample Design

The 2016 ACS Content Test consisted of a nationally representative sample of 70,000 residential addresses in the United States, independent of the production ACS sample. The Content Test sample universe did not include GQs, nor did it include housing units in Alaska, Hawaii, or Puerto Rico.[9] The sample design for the Content Test was largely based on the ACS production sample design with some modifications to better meet the test objectives.[10] The modifications included adding an additional level of stratification by stratifying addresses into high and low self-response areas, oversampling addresses from low self-response areas to ensure equal response from both strata, and sampling units as pairs.[11] The high and low self-response strata were defined based on ACS self-response rates at the tract level. Sampled pairs were formed by first systematically sampling an address within the defined sampling stratum and then pairing that address with the address listed next in the geographically sorted list. Note that the pair was likely not neighboring addresses. One member of the pair was randomly assigned to receive the control version of the question and the other member was assigned to receive the test version of the question, thus resulting in a sample of 35,000 control cases and 35,000 test cases.

As in the production ACS, if efforts to obtain a response by mail or telephone were unsuccessful, attempts were made to interview in person a sample of the remaining nonresponding addresses (see Section 2.2 Data Collection for more details). Addresses were sampled at a rate of 1-in-3, with some exceptions that were sampled at a higher rate.[12] For the Content Test, the development of workload estimates for CATI and CAPI did not take into account the oversampling of low response areas. This oversampling resulted in a higher than expected workload for CATI and CAPI and therefore required more budget than was allocated. To address this issue, the CAPI sampling rate for the Content Test was adjusted to meet the budget constraint.

[9] Alaska and Hawaii were excluded for cost reasons. GQs and Puerto Rico were excluded because the sample sizes required to produce reliable estimates would be overly large and burdensome, as well as costly.

[10] The ACS production sample design is described in Chapter 4 of the ACS Design and Methodology report (U.S. Census Bureau, 2014).

[11] Tracts with the highest response rate based on data from the 2013 and 2014 ACS were assigned to the high response stratum in such a way that 75 percent of the housing units in the population (based on 2010 Census estimates) were in the high response areas; all other tracts were designated in the low response strata. Self-response rates were used as a proxy for overall cooperation. Oversampling in low response areas helps to mitigate larger variances due to CAPI subsampling. This stratification at the tract level was successfully used in previous ACS Content Tests, as well as the ACS Voluntary Test in 2003.

[12] The ACS production sample design for CAPI follow-up is described in Chapter 4, Section 4.4 of the ACS Design and Methodology report (U.S. Census Bureau, 2014).

2.2 Data Collection

The field test occurred in parallel with the data collection activities for the March 2016 ACS production panel, using the same basic data collection protocol as production ACS with a few differences as noted below. The data collection protocol consisted of three main data collection operations: 1) a six-week mailout period, during which the majority of internet and mailback responses were received; 2) a one-month CATI period for nonresponse follow-up; and 3) a one-month CAPI period for a sample of the remaining nonresponse. Internet and mailback responses were accepted until three days after the end of the CAPI month.

As indicated earlier, housing units included in the Content Test sample were randomly assigned to a control or test version of the questions. CATI interviewers were not assigned specific cases; rather, they worked the next available case to be called and therefore conducted interviews for both control and test cases. CAPI interviewers were assigned Content Test cases based on their geographic proximity to the cases and therefore could also conduct both control and test cases.

The ACS Content Test's data collection protocol differed from the production ACS in a few significant ways. The Content Test analysis did not include data collected via the Telephone Questionnaire Assistance (TQA) program since those who responded via TQA used the ACS production TQA instrument. The Content Test excluded the telephone Failed Edit Follow-Up (FEFU) operation.[13] Furthermore, the Content Test had an additional telephone reinterview operation used to measure response reliability. We refer to this telephone reinterview component as the Content Follow-Up, or CFU. The CFU is described in more detail in Section 2.3.

ACS production provides Spanish-language versions of the internet, CATI, and CAPI instruments, and callers to the TQA number can request to respond in Spanish, Russian, Vietnamese, Korean, or Chinese. The Content Test had Spanish-language automated instruments; however, there were no paper versions of the Content Test questionnaires in Spanish.[14] Any case in the Content Test sample that completed a Spanish-language internet, CATI, or CAPI response was included in analysis. However, if a case sampled for the Content Test called TQA to complete an interview in Spanish or any other language, the production interview was conducted and the response was excluded from the Content Test analysis. This was due to the low volume of non-English language cases and the operational complexity of translating and implementing several language instruments for the Content Test. CFU interviews for the Content Test were conducted in either Spanish or English. The practical need to limit the language response options for Content Test respondents is a limitation to the research, as some respondents self-selected out of the test.

[13] In ACS production, paper questionnaires with an indication that there are more than five people in the household or questions about the number of people in the household, and self-response returns that are identified as being vacant or a business or lacking minimal data are included in FEFU. FEFU interviewers call these households to obtain any information the respondent did not provide.

[14] In the 2014 ACS, respondents requested 1,238 Spanish paper questionnaires, of which 769 were mailed back. From that information, we projected that fewer than 25 Spanish questionnaires would be requested in the Content Test.

2.3 Content Follow-Up

For housing units that completed the original interview, a CFU telephone reinterview was also conducted to measure response error.[15] A comparison of the original interview responses and the CFU reinterview responses was used to answer research questions about response error and response reliability.

A CFU reinterview was attempted with every household that completed an original interview for which there was a telephone number. A reinterview was conducted no sooner than two weeks (14 calendar days) after the original interview. Once the case was sent to CFU, it was to be completed within three weeks. This timing balanced two competing interests: (1) conducting the reinterview as soon as possible after the original interview to minimize changes in truth between the two interviews, and (2) not making the two interviews so close together that the respondents were simply recalling their previous answers. Interviewers made two call attempts to interview the household member who originally responded, but if that was not possible, the CFU reinterview was conducted with any other eligible household member (15 years or older).

The CFU asked basic demographic questions and a subset of housing and detailed person questions that included all of the topics being tested, with the exception of Telephone Service, and any questions necessary for context and interview flow to set up the questions being tested. [16] All CFU questions were asked in the reinterview, regardless of whether or not a particular question was answered in the original interview. Because the CFU interview was conducted via telephone, the wording of the questions in CFU followed the same format as the CATI nonresponse interviews. Housing units assigned to the control version of the questions in the original interview were asked the control version of the questions in CFU; housing units assigned to the test version of the questions in the original interview were asked the test version of the questions in CFU. The only exception was for retirement, survivor, and disability income, for which a different set of questions was asked in CFU. [17]

2.4 Industry and Occupation Coding Procedures

Much of the analysis in this report used the coded responses, which are assigned based on the write-in responses. There are two write-in fields each for the questions on Industry and Occupation (images of the test and control versions are presented in section 1.3). For Industry, the questions on Employer Name and Kind of Business include write-in fields. For Occupation, the questions on Job Title and Job Duties include write-in fields.

The coding process assigns one of 269 Census Industry categories and one of 539 Census Occupation categories, including Military, to the write-in responses provided for the Industry and

[15] Throughout this report the "original interview" refers to responses completed via paper questionnaire, internet, CATI, or CAPI.

[16] Because the CFU interview was conducted via telephone the Telephone Service question was not asked. We assume that CFU respondents have telephone service.

[17] Refer to the 2016 ACS Content Test Report on Retirement Income for a discussion on CFU questions for Survivor, Disability, and Retirement Income.

Occupation questions.[18] The Census categories are based on the North American Industry Classification System (NAICS) and the SOC, but are less detailed.[19] Industry and occupation codes are 4-digit codes. Beginning in 2012, an automated process (i.e., autocoder) was implemented to code the responses to supplement clerical coding. Both the autocoder and clerical coding processes use the following variables to interpret the text write-ins and assign categories for Industry and Occupation:

- State and County of Residence
- Age
- Sex
- Educational Attainment
- Class of Worker
- Active Duty Checkbox
- Employer Name (write-in)
- Kind of Business (write-in)
- Industry Type Checkbox
- Job Title (write-in)
- Job Duties (write-in)

In production, coding begins with the automated coding of industry and occupation. Cases not fully coded by the autocoder are transferred to the National Processing Center's (NPC) Industry and Occupation Coding Units, divided into batches of 100 cases, and then clerically coded.[20] Cases that cannot be clerically coded at this step are then sent to expert coders with more training and resources, internally known as Referralists. Coders are allowed to change Class of Worker during coding so that value may differ on the input coding file.

The autocoder was not used in the Content Test industry and occupation coding process.[21] All cases were coded manually by clerical coders, then by Referralists if needed. As in production, coders were able to change the Class of Worker value during coding for the test and control treatments.[22] During regular production coding, any industry and occupation data that cannot be

[18] For more information on Census industry and occupation codes, see http://www.census.gov/people/io/methodology/.

[19] For more information on the North American Industry Classification System (NAICS), see http://www.census.gov/eos/www/naics/.

[20] For ACS, maximum batch sizes have 100 cases. Usually, the last batch in a coding file may contain less than 100 cases.

[21] The autocoder uses dictionaries that contain words or phrases commonly found in the ACS industry and occupation write-ins and the specific industry and occupation codes that they are most commonly associated with. A regression model is used to select the "best" code from among the list of possible codes. Data from the 2010 ACS survey year were used in creating the autocoder dictionaries and regression models required for automated coding. Since the Industry and Occupation questions and the number of characters allowed were different between the control and test versions, it was determined that using the autocoder for the test treatment would not be appropriate. Also, we expected the autocoder to code the control treatment cases at a higher rate since the autocoder dictionaries are based on the same Industry and Occupation questions used on the control treatment. This would have had an effect on the interim coding referral rates and median coding time.

[22] During the 2016 ACS Content Test, the Class of Worker and the Industry and Occupation questions were tested concurrently; results are presented separately.

coded are assigned or imputed a code in post-processing editing. The data from the Content Test were not subjected to the post-processing editing.

2.5 Analysis Metrics

This section describes the metrics used to assess the revised versions of the Industry and Occupation questions. These metrics include item missing data rates, response distributions, response error, and other metrics. This section also describes the methodology used to calculate unit response rates and standard errors for the test.

All Content Test data were analyzed without imputation due to our interest in how question changes or differences between versions of new questions affected "raw" responses, not the final edited variables. Some editing of responses was done for analysis purposes, such as collapsing response categories or modes together or calculating a person's age based on his or her date of birth.

All estimates from the ACS Content Test were weighted. Analysis involving data from the original interviews used the final weights that take into account the initial probability of selection (the base weight) and CAPI subsampling. For analysis involving data from the CFU interviews, the final weights were adjusted for CFU nonresponse to create CFU final weights.

The significance level for all hypothesis tests is $\alpha = 0.1$. Since we are conducting numerous comparisons between the control and test treatments, there is a concern about incorrectly rejecting a hypothesis that is actually true (a "false positive" or Type I error). The overall Type I error rate is called the familywise error rate and is the probability of making one or more Type I errors among all hypotheses tested simultaneously. When adjusting for multiple comparisons, the Holm-Bonferroni method was used (Holm, 1979).

2.5.1 Unit Response Rates and Demographic Profile of Responding Households

The unit response rate is generally defined as the proportion of sample addresses eligible to respond that provided a complete or sufficient partial response.[23] Unit response rates from the original interview are an important measure to look at when considering the analyses in this report that compare responses between the control and test versions of the survey questionnaire. High unit response rates are important in mitigating potential nonresponse bias.

For both control and test treatments, we calculated the overall unit response rate (all modes of data collection combined) and unit response rates by mode: internet, mail, CATI, and CAPI. We also calculated the total self-response rate by combining internet and mail modes together. Some Content Test analyses focused on the different data collection modes for topic-specific evaluations, thus we felt it was important to include each mode in the response rates section. In addition to those rates, we calculated the response rates for high and low response areas because analysis for some Content Test topics was done by high and low response areas. Using the

[23] A response is deemed a "sufficient partial" when the respondent gets to the first question in the detailed person questions section for the first person in the household.

Census Bureau's Planning Database (U.S. Census Bureau, 2016), we defined these areas at the tract level based on the low response score.

The universe for the overall unit response rates consists of all addresses in the initial sample (70,000 addresses) that were eligible to respond to the survey. Some examples of addresses ineligible for the survey were a demolished home, a home under construction, a house or trailer that was relocated, or an address determined to be a permanent business or storage facility. The universe for self-response (internet and mail) rates consists of all mailable addresses that were eligible to respond to the survey. The universe for the CATI response rate consists of all nonrespondents at the end of the mailout month from the initial survey sample that were eligible to respond to the survey and for whom we possessed a telephone number. The universe for the CAPI response rates consists of a subsample of all remaining nonrespondents (after CATI) from the initial sample that were eligible to respond to the survey. Any nonresponding addresses that were sampled out of CAPI were not included in any of the response rate calculations.

We also calculated the CFU interview unit response rate overall and by mode of data collection of the original interview and compared the control and test treatments because response error analysis (discussed in Section 2.5.6.) relies upon CFU interview data. Statistical differences between CFU response rates for control and test treatments will not be taken as evidence that one version is better than the other. For the CFU response rates, the universe for each mode consists of housing units that responded to the original questionnaire in the given mode (internet, mail, CATI, or CAPI) and were eligible for the CFU interview. We expected the response rates to be similar between treatments; however, we calculated the rates to verify that assumption.

Another important measure to look at in comparing experimental treatments is the demographic profile of the responding households in each treatment. The Content Test sample was designed with the intention of having respondents in both control and test treatments exhibit similar distributions of socioeconomic and demographic characteristics. Similar distributions allow us to compare the treatments and conclude that any differences are due to the experimental treatment instead of underlying demographic differences. Thus, we analyzed distributions for data from the following response categories: *age, sex, educational attainment,* and *tenure.* The topics of *race, Hispanic origin,* and *relationship* are also typically used for demographic analysis; however, those questions were modified as part of the Content Test, so we could not include them in the demographic profile. Additionally, we calculated *average household size* and the *language of response* for the original interview.[24]

For response distributions, we used chi-square tests of independence to determine statistical differences between control and test treatments. If the distributions were significantly different, we performed additional testing on the differences for each response category. To control for the overall Type I error rate for a set of hypotheses tested simultaneously, we performed multiple-comparison procedures with the Holm-Bonferroni method (Holm, 1979). A family for our response distribution analysis was the set of p-values for the overall characteristic categories (*age, sex, educational attainment,* and *tenure*) and the set of p-values for a characteristic's response categories if the response distributions were found to have statistically significant

[24] Language of response analysis excludes paper questionnaire returns because there was only an English questionnaire.

differences. To determine statistical differences for *average household size* and the *language of response* of the original interview we performed two-tailed hypothesis tests.

For all response-related calculations mentioned in this section, addresses that were either sampled out of the CAPI data collection operation or that were deemed ineligible for the survey were not included in any of the universes for calculations. Unmailable addresses were also excluded from the self-response universe. For all unit response rate estimates, differences, and demographic response analysis, we used replicate base weights adjusted for CAPI sampling (but not adjusted for CFU nonresponse).

2.5.2 Item Missing Data Rates

Respondents leave items blank for a variety of reasons including not understanding the question (clarity), their unwillingness to answer a question as presented (sensitivity), and their lack of knowledge of the data needed to answer the question. The item missing data rate (for a given item) is the proportion of eligible units, housing units for household-level items or persons for person-level items, for which a required response (based on skip patterns) is missing.

We calculated an item missing rate for Industry and for Occupation across and within each mode of data collection (internet, mail, CATI, and CAPI). The universe was comprised of person-level records for persons who were 15 years of age or older who met the following criteria:

- Worked last week for pay at a job or business.
- Performed any work last week for pay, even for as little as one hour.
- Last worked within the last 12 months or one to five years ago.

The time period for work is relative to when the respondent completed questionnaire.

The Industry item was classified as missing for a person-level record if both of the Industry write-in fields were blank. The Occupation item was classified as missing for a person-level record if both of the Occupation write-in fields were blank.

We compared the item missing data rates between the control and test versions of the questions via a two-tailed t-test.

2.5.3 Codeable Rates

The codeable rate is the proportion of Industry/Occupation person-level records in universe that were assigned a valid industry/occupation code. The universe for Industry was the subset of the universe defined in Section 2.5.2 for which the assigned industry code was valid and nonblank. The universe for Occupation was the subset of this same universe for which the assigned occupation code was valid and nonblank. We calculated codeable rates for Industry and Occupation across and within data collection mode (internet, mail, CATI, and CAPI).

For this study, the industry codes were categorized into the 21 four-digit industry categories shown in Table 1. The occupation codes were categorized into the 23 four-digit Occupational

categories shown in Table 2. For Industry and for Occupation, we compared the codeable rates of the control and test versions of the questions via a two-tailed t-test.

Table 1. Census Industry Codes (corresponding to NAICS* Industry sectors)

Description of NAICS Industry Sector	Range of Census Industry Codes
Agriculture, forestry, fishing and hunting	0170-0290
Mining quarrying, and oil and gas extraction	0370-0490
Construction	0770
Manufacturing	1070-3990
Wholesale trade	4070-4590
Retail trade	4670-5790
Transportation and warehousing	6070-6390
Utilities	0570-0690
Information	6470-6780
Finance and insurance	6870-6990
Real estate and rental and leasing	7070-7190
Professional, scientific, and technical services	7270-7490
Management of companies and enterprises	7570
Administrative and support and waste management services	7580-7790
Educational services	7860-7890
Health care and social assistance	7970-8470
Arts, entertainment, and recreation	8560-8590
Accommodation and food services	8660-8690
Other public services, except public administration	8770-9290
Public administration	9370-9590
Military	9670-9870

Source: U.S. Census Bureau, 2012 Census Industry Code List. For more information on Census Industry codes, see: http://www.census.gov/people/io/methodology/.

* North American Industry Classification System (NAICS)

Table 2. Census Occupation Codes (corresponding to SOC* major groups)

Description of SOC Major Group	Range of Census Occupation Codes
Management occupations	0010-0430
Business and financial operations occupations	0500-0950
Computer and mathematical occupations	1000-1240
Architecture and engineering occupations	1300-1560
Life, physical, and social science occupations	1600-1965
Community and social services occupations	2000-2060
Legal occupations	2100-2160
Education, training, and library occupations	2200-2550
Arts, design, entertainment, sports, and media occupations	2600-2960
Healthcare practitioner and technical occupations	3000-3540
Healthcare support occupations	3600-3655
Protective service occupations	3700-3955
Food preparation and serving related occupations	4000-4160
Building and grounds cleaning and maintenance occupations	4200-4250
Personal care and service occupations	4300-4650
Sales and related occupations	4700-4965
Office and administrative support occupations	5000-5940
Farming, fishing, and forestry occupations	6005-6130
Construction and extraction occupations	6200-6940
Installation, maintenance, and repair occupations	7000-7630
Production occupations	7700-8965
Transportation and material moving occupations	9000-9750
Military specific occupations	9800-9830

Source: U.S. Census Bureau, 2010 Census Occupation Code List. For more information on Census occupation codes, see: http://www.census.gov/people/io/methodology/.
* Standard Occupational Classification

2.5.4 Interim Referral Rates

The Industry and Occupation autocoder was not used for the Content Test. The coding was performed entirely by clerical coders. Cases that could not be coded at the first step by the clerical coders were sent to Referralist coders. Referralist coders are expert coders who have more training and resources than clerical coders. Examples of situations that require referring of the case include: (1) not enough information is available within the case; (2) restrictions are listed in one of the coding indexes; (3) procedural rules define when to refer a certain situation; and (4) inconsistency exists between the various responses in the case. This two-step process increases coding efficiency.

Referring a case does not necessarily imply a clerical coder was unsuccessful. Referralist coders have access to additional resources that allow them to determine the more accurate code. These materials include the NAICS and SOC Manual and internet websites. By design, the Industry and Occupation data were not imputed in post-processing editing for this test in order to evaluate how question changes affect actual "raw" responses, not the final edited variables.

We combined the results for Industry and Occupation, since if either one of these codes needs referral, the whole case has to be referred. When this happens, any of the two codes can be

changed by the Referralist. The interim referral rate is the number of cases referred divided by the total number of cases coded. We compared the interim referral rates of the control and test versions of the questions via a two-tailed t-test.

2.5.5 Response Distributions

Comparing the response distributions between the control version of a question and the test version of a question allows us to assess whether the question change affected the resulting estimates. Comparisons were made using Rao-Scott chi-squared tests (Rao & Scott, 1987) for distribution and t-tests for single categories when the corresponding distributions were found to be statistically different. Proportion estimates were calculated as:

$$\text{Category proportion} = \frac{\text{weighted count of valid responses in category}}{\text{weighted count of all valid responses}}$$

The categories examined for Industry and Occupation were shown in section 2.5.3 in Tables 1 and 2, respectively.

2.5.6 Response Error

Response error occurs for a variety of reasons, such as flaws in the survey design, misunderstanding of the questions, misreporting by respondents, or interviewer effects. There are two components of response error: response bias and simple response variance. Response bias is the degree to which respondents consistently answer a question incorrectly. Simple response variance is the degree to which respondents answer a question inconsistently. A question has good response reliability if respondents tend to answer the question consistently. Re-asking the same question of the same respondent (or housing unit) allows us to measure response variance.

We measured simple response variance by comparing valid responses to the CFU reinterview with valid responses to the corresponding original interview.[25] The Census Bureau has frequently used content reinterview surveys to measure simple response variance for large demographic data collection efforts, including the 2010 ACS Content Test, and the 1990, 2000, and 2010 decennial censuses (Dusch & Meier, 2012).

The following measures were used to evaluate consistency:

- 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.

[25] A majority of the CFU interviews were conducted with the same respondent as the original interview (see the Limitations section for more information).

The GDR, and subsequently the simple response variance, are calculated using the following table and formula.

Table 3. Interview and Reinterview Counts for Each Response Category Used for Calculating the Gross Difference Rate and Index of Inconsistency

	Original Interview "Yes"	Original Interview "No"	**Reinterview Totals**
CFU Reinterview "Yes"	a	b	**a + b**
CFU Reinterview "No"	c	d	**c + d**
Original Interview Totals	**a + c**	**b + d**	**n**

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

a = weighted count of units in the category of interest for both the original interview and reinterview
b = weighted count of units NOT in the category of interest for the original interview, but in the category for the reinterview
c = weighted count of units in the category of interest for the original interview, but NOT in the category for the reinterview
d = weighted count of units NOT in the category of interest for either the original interview or the reinterview
n = total units in the universe = a + b + c + d.

The GDR for a specific response category is the percent of inconsistent answers between the original interview and the reinterview (CFU). We calculate the GDR for a response category as

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

Statistical significance between the GDR for a specific response category between the control and test treatments is determined using a one-tailed t-test.

In order 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 (Flanagan, 1996).

SV is the part of total variance resulting from the differences among 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 the total variance is due to inconsistency in responses, as measured by SRV and is calculated as:

$$\text{IOI} = \frac{n(b+c)}{(a+c)(c+d)+(a+b)(b+d)} \times 100$$

Per the Census Bureau's general rule, index values of less than 20 percent indicate low inconsistency, 20 to 50 percent indicate moderate inconsistency, and over 50 percent indicate high inconsistency.

An IOI is computed for each response category and an overall index of inconsistency, called the L-fold index of inconsistency, is reported for the entire distribution. The L-fold index is a weighted average of the individual indexes computed for each response category.

When the sample size is small, the reliability estimates are unstable. Therefore, we do not report the IOI and GDR values for categories with a small sample size, as determined by the following formulas: $2a + b + c < 40$ or $2d + b + c < 40$, where a, b, c, and d are unweighted counts as shown in Table 3 above (see Flanagan 1996, p. 15).

The measures of response error assume that those characteristics in question did not change between the original interview and the CFU interview. To the extent that this assumption is incorrect, we assume that it is incorrect at similar rates between the control and test treatments.

In calculating the IOI reliability measures, the assumption is that the expected value of the error in the original interview is the same as in the CFU reinterview. This assumption of parallel measures is necessary for the SRV and IOI to be valid. In calculating the IOI measures for this report, we found this assumption was not met for the response categories specified in the limitations section (see Section 4).

Biemer (2011, pp. 56-58) provides an example where the assumption of parallel measures is not met, but does not provide definitive guidelines for addressing it. In Biemer's concluding remarks, he states, "...both estimates of reliability are biased to some extent because of the failure of the parallel assumptions to hold." Flanagan (2001) addresses this bias problem and offers the following adjustment to the IOI formula:

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

This formula was tested on selected topics in the 2016 ACS Content Test. The $\text{IOI}_{\text{testimate}}$ resulted in negligible reduction in the IOI values. For this reason, we did not recalculate the IOI values using $\text{IOI}_{\text{testimate}}$. Similar to Biemer (2011, p. 58), we acknowledge that for some cases, the estimate of reliability is biased to some extent.

2.5.7 Analysis of Industry and Occupation Write-in Fields

One goal of this test was to obtain more detailed write-in responses, especially for Occupation, in order to make the coding process easier and more accurate, and to investigate if more code refinement is possible in the future. The Census Bureau currently aggregates codes within both occupation (SOC) and industry (NAICS). However, we would like to provide more detail if possible. At present, our ability to provide detail is limited by the amount of detail provided by respondents. Specificity of responses for each of the four write-ins is an important factor in comparing the test and control treatments. We examined mean character count, mean word count, specificity, and median coding time.

Character and word counts measured objectively if more detail was provided by respondents. We compared each of the four Industry and Occupation write-in fields independently. Of particular interest was the Job Duties write-in field, which was expanded from 60 to 100 characters in the test version and is visually cued in the internet and mail modes. We expected character and word counts to be higher for test than for control for all write-ins, with the exception of Employer Name.

An additional indicator of the specificity of write-in responses was measured from a qualitative review conducted by Referralists. The Referralists reviewed test and control records of selected detailed Occupation categories that are the hardest to get respondents to provide more accuracy on by modes of collection. Then, they compared the content (detail and quality) of the write-in responses. For both Industry and Occupation, expert coders scored the test records on the level of detail, compared them to the control records, and determined if the test records would result in better, more precise codes using the current code categories.

The changes to the wording of the Industry and Occupation questions and the increase in the number of characters allowed for the Job Duties question were intended to help respondents better understand what information we are asking for. We expected that their written responses would be easier to code, as determined by coding time. We expected coding time to be lower in the test version.

2.5.7.1 Mean Character Count

The *mean character count* is the average number of characters provided in each of the four Industry or Occupation write-in fields. Characters include letters, spaces, question marks, exclamation points, numbers, etc. We performed a one-tail t-test to test if the mean character count for the test version of the question was greater than that of the control version.

2.5.7.2 Mean Word Count

The *mean word count* is the average number of words provided in each of the four Industry or Occupation write-in fields. We performed a one-tail t-test to test if the mean word count for the test version of the question was greater than that of the control version.

2.5.7.3 Specificity

The objective of the expert, qualitative coding review was to compare the detail and quality of the content of the Industry and Occupation write-ins from the control and test treatments for selected Occupation categories by mode of collection and code categories. The review included a total of 6,974 cases in 39 different occupation code categories, as well as *Active Duty Military* and *National Guard*.[26] Referralists received control cases, test cases, and guide questions for their review.

The cases were classified by code and mode of data collection. They contained the following information:

1. Control or Test Flag[27]
2. Mode of Data Collection
3. Record Identification Number
4. Age
5. Sex
6. Employer Name (write-in)
7. Kind of Business (write-in)
8. Industry Type Checkbox
9. Job Title (write-in)
10. Job Duties (write-in)
11. Industry Code
12. Occupation Code
13. Class of Worker (pre-coding)[28]
14. Class of Worker (post-coding)[29]
15. Active Duty Flag[30]

In conducting this comparison, Referralists were guided by eight questions. These questions, listed below, referred to the code category as a whole and not to each case. For the first two questions, Referralists were required to measure the comparison with a scale from 1 to 5, where 1 has the least detail and 5 the most detail.

On a scale from 1 to 5 (1=less detail, 3=neutral, and 5=more detail):

- In comparison to the control version, is there more detail or less detail on the Industry test write-ins?

[26] For full list of code categories, see C-1 in Appendix C.
[27] This flags the case as part of the control treatment or test treatment.
[28] This is the value for the Class of Worker variable before the file was sent to the coding clerks.
[29] This is the value for the Class of Worker variable after the file was sent to the coding clerks. Industry and Occupation coding clerks are allowed to change the value for Class of Worker based on the Industry and Occupation information in each case.
[30] When a respondent states that he or she was *Active Duty Military*, then the case was flagged as *Active Duty*.

- In comparison to the control version, is there more detail or less detail on the Occupation test write-ins?

For the next two questions, Referralists were required to write a yes or no response:

- In comparison to the control version, would the amount of detail in the test version result in a better, more precise industry code?
- In comparison to the control version, would the amount of detail in the test version result in a better, more precise occupation code?

The last four items required a brief, but thorough, explanation of the Referralist evaluation. The Referralists were expected to compare the variables on the test and control cases by code category and mode of collection. The questions were:

Considering full coding information over all cases for this code category and mode, comparing control and test:

- Is there a specific piece of information missing that would help code this industry?
- Is there a specific piece of information missing that would help code this occupation?
- Is there a critical element that helped code industry?
- Is there a critical element that helped code occupation?

2.5.8 Median Coding Time

We initially intended to assess if the mean coding time for the test version of the question was lower than the control. However, due to the manner in which the coding times were collected, the data were highly skewed and the recorded mean would not provide an accurate measure, therefore we computed the median coding time. To test for differences, we used a non-parametric test called the Wilcoxon Rank Sum Test.[31] The coding times were based on the time it took a clerical coder to code a batch of 100 CFU cases.[32] The coding time (in minutes) was calculated from the time coding for a batch started until it finished. There were a total of 16 clerical coders coding 126 CFU batches for the control treatment and 124 CFU batches for the test treatment.

2.5.9 Benchmarks

We initially intended to benchmark the responses to the Occupation question in the 2016 ACS Content Test to information from the Bureau of Labor Statistics' Occupational Employment Statistics (OES) program. Our objective was to compare occupations that historically have been difficult to code to a greater level of detail. This comparison would have allowed us to tell whether our results were grossly different from another reliable resource.

However, it was determined that benchmarking the 2016 ACS Content Test to the May 2015 OES program would not have been appropriate because of differences in the survey

[31] For more information on the Wilcoxon Rank Sum Test, see http://analytics.ncsu.edu/sesug/2004/TU04-Pappas.pdf.

[32] For ACS, maximum batch sizes have 100 cases. Usually, the last batch in a coding file may contain less than 100 cases.

methodology. The OES program conducts a semi-annual mail survey designed to produce estimates of employment and wages for specific occupations. It collects data for the payroll period, including the 12th day of May or November, on wage and salary workers in nonfarm establishments to produce employment and wage estimates for about 800 occupations. More importantly, the OES program does not cover the self-employed, owners and partners in unincorporated firms, household workers, or unpaid family workers in its survey sample. The workforces represented in the OES program and the 2016 ACS Content Test are different, thus benchmarking would have been inappropriate.

2.5.10 Industry and Class of Worker Consistency

The Class of Worker question categorizes workers according to the type of ownership of the employing organization. It is important that a worker's Class of Worker category and industry code be consistent. For example, due to certain NAICS restrictions, anyone with industry code in public administration must have a Class of Worker category of government worker. For this analysis, we assumed for both test and control that respondents report Class of Worker, Industry and Occupation, and Labor Force data for the same job, or that they at least equally misreported across variables these data. We tested the differences in the consistency rates between the control and test treatments using two-tailed t-tests. We compared the response for the unedited Class of Worker value going into the coding process with the clerk code on unedited industry.[33] We also conducted a comparison of a sample from selected categories of Class of Worker with the two Industry write-ins, Employer Name and Kind of Business, based on our current post-processing editing procedures. For example, in these analyses, we checked to see if individuals whose response to Class of Worker was Government selected *Public Administration* as the industry. For those that reported *Active Duty Military* for the Class of Worker question, we checked to see if they selected *Military* for industry. Additionally, we evaluate how frequently those who selected Private sector for Class of Worker selected *Public Administration* or *Military* for industry, which would be an incorrect response option if the responses were about the same job. For a subset of inconsistent records, we intended to compare the coding output from the unedited Class of Worker responses with unedited industry responses to see if the clerks were able to correct the Class of Worker response to be consistent with other information on the record. Due to time restrictions, this evaluation could not be completed.

2.5.11 Standard Error Calculations

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, 2014, Chapter 12). We calculated the variance for each rate and difference using the formula below. The standard error of the estimate (X_0) is the square root of the variance.

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

[33] Table C-3 Class of Worker and Industry Consistency Check can be found in Appendix C.

where:

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

3 DECISION CRITERIA FOR INDUSTRY AND OCCUPATION

Before fielding the 2016 ACS Content Test, we identified which of the metrics would be given higher importance in determining which version of the question would be recommended for inclusion in the ACS moving forward. Table 4 identifies the research questions and associated metrics in priority order.

Table 4. Decision Criteria for the topic of Industry and Occupation

Research Questions	Decision Criteria in order of priority
1, 2, 3, 4	▪ The item missing data rates for Industry and Occupation for the test version should be the same or lower than the control version, for all modes. ▪ The item codeable rates for Industry and Occupation for the test version should be the same or higher than the control version, for all modes.
10	▪ For Industry and Occupation, the test version should have the same or higher qualitative scores on specificity of responses than the control version, for all modes.
5	▪ The interim referral rates for the test version should be the same or lower than for the control version.
11, 12	▪ For Industry and Occupation, the test version should have the same or higher mean word count and mean character count than the control version, for all modes.
8, 9	▪ The response reliability for Industry and Occupation should be the same or higher for the test version than the control version.
14	▪ The proportion of workers with consistent Industry and Class of Worker should be the same or higher for test than for control.
6, 7	▪ The Industry and Occupation response distributions between the test and control versions should have no differences.
13	▪ The coder median coding time for test should be less than control.

4 LIMITATIONS

CATI and CAPI interviewers were assigned control and test treatment cases, as well as production cases. The potential risk of this approach is the introduction of a cross-contamination or carry-over effect due to the same interviewer administering multiple versions of the same question item. Interviewers are trained to read the questions verbatim to minimize this risk, but there still exists the possibility that an interviewer may deviate from the scripted wording of one question version to another. This could potentially mask a treatment effect from the data collected.

Interviews were only conducted in English and Spanish. Respondents who needed language assistance in another language were not able to participate in the test. Additionally, the 2016 ACS Content Test was not conducted in Alaska, Hawaii, or Puerto Rico. Any conclusions drawn from this test may not apply to these areas or populations.

For statistical analysis specific to the mail mode, there may be bias in the results because of unexplained unit response rate differences between the control and test treatments.

We were not able to conduct demographic analysis by relationship status, race, or ethnicity because these topics were tested as part of the Content Test.

The CFU reinterview was not conducted in the same mode of data collection for households that responded by internet, by mail, or by CAPI in the original interview since CFU interviews were only administered using a CATI mode of data collection. As a result, the data quality measures derived from the reinterview may include some bias due to the differences in mode of data collection.

To be eligible for a CFU reinterview, respondents needed to either provide a telephone number in the original interview or have a telephone number available to the Census Bureau through reverse address look up. As a result, 2,284 of the responding households (11.8 percent with a standard error of 0.2) from the original control interviews and 2,402 of the responding households (12.4 percent with a standard error of 0.2) from the original test interviews were not eligible for the CFU reinterview. The difference between the control and test treatments is statistically significant (p-value=0.06).

Although we reinterviewed the same person who responded in the original interview when possible, we interviewed a different member of the household in the CFU for 7.5 percent (standard error of 0.4) of the CFU cases for the control treatment and 8.4 percent (standard error of 0.5) of the CFU cases for the test treatment.[34] The difference between the test and control treatments is not statistically significant (p-value=0.26). This means that differences in results between the original interview and the CFU for these cases could be due in part to having different people answering the questions. However, those changes were not statistically significant between the control and test treatments and should not impact the conclusions drawn from the reinterview.

The 2016 ACS Content Test does not include the production weighting adjustments for seasonal variations in ACS response patterns, nonresponse bias, and under-coverage bias. As a result, any estimates derived from the 2016 ACS Content Test data do not provide the same level of inference as the production ACS and cannot be compared to production estimates.

In developing initial workload estimates for CATI and CAPI, we did not take into account the fact that we oversampled low response areas as part of the Content Test sample design. Therefore, workload and budget estimates were too low. In order to stay within budget, the CAPI

[34] This is based on comparing the first name of the respondent between the original interview and the CFU interview. Due to a data issue, we were not able to use the full name to compare.

workload was subsampled more than originally planned. This caused an increase in the variances for the analysis metrics used.

An error in addressing and assembling the materials for the 2016 ACS Content Test caused some Content Test cases to be mailed production ACS questionnaires instead of Content Test questionnaires. There were 49 of these cases that returned completed questionnaires, and they were all from the test treatment. These cases were excluded from the analysis. Given the small number of cases affected by this error, there is very little effect on the results.

Questionnaire returns were expected to be processed and keyed within two weeks of receipt. Unfortunately, a check-in and keying backlog prevented this requirement from being met, thereby delaying eligible cases from being sent to CFU on a schedule similar to the other modes. Additionally, the control treatment questionnaires were processed more quickly in keying than the test treatment questionnaires resulting in a longer delay for test mail cases to be eligible for CFU. On average, it took 18 days for control cases to become eligible for CFU; it took 20 days for test cases. The difference is statistically significant. This has the potential to impact the response reliability results.

For the expert qualitative coding review, the respondent's educational attainment was not provided to the Referralist coders for either treatment. Educational attainment is information normally used during the coding of Industry and Occupation responses.

We were unable to perform benchmark comparisons due to differences in data collection and universes between the 2015 May OES and the ACS.

5 RESEARCH QUESTIONS AND RESULTS

This section presents the results from the analyses of the 2016 ACS Content Test data for Industry and Occupation questions. An analysis of unit response rates is presented first followed by topic-specific analyses. For the topic-specific analyses, each research question is restated, followed by corresponding data and a brief summary of the results.

5.1 Unit Response Rates and Demographic Profile of Responding Households

This section provides results for unit response rates for both control and test treatments for the original Content Test interview and for the CFU interview. It also provides results of a comparison of socioeconomic and demographic characteristics of respondents in both control and test treatments.

5.1.1 Unit Response Rates for the Original Content Test Interview

The unit response rate is generally defined as the proportion of sample addresses eligible to respond that provided a complete or sufficient partial response. We did not expect the unit response rates to differ between treatments. This is important because the number of unit responses should also affect the number of item responses we receive for analyses done on specific questions on the survey. Similar item response universe sizes allow us to compare the

treatments and conclude that any differences are due to the experimental treatment instead of differences in the populations sampled for each treatment.

Table 5 shows the unit response rates for the original interview for each mode of data collection (internet, mail, CATI, and CAPI), all modes combined, and both self-response modes (internet and mail combined) for the control and test treatments. When looking at the overall unit response rate (all modes combined) the difference between control (93.5 percent) and test (93.5 percent) is less than 0.1 percentage points and is not statistically significant.

Table 5. Original Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode

Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
All Modes	19,400	93.5 (0.3)	19,455	93.5 (0.3)	<0.1 (0.4)	0.98
Self-Response	13,131	52.9 (0.5)	13,284	53.7 (0.5)	-0.8 (0.6)	0.23
Internet	8,168	34.4 (0.4)	8,112	34.1 (0.4)	0.4 (0.6)	0.49
Mail	4,963	18.4 (0.3)	5,172	19.6 (0.3)	-1.2 (0.5)	0.01*
CATI	872	8.7 (0.4)	880	9.2 (0.4)	-0.4 (0.6)	0.44
CAPI	5,397	83.5 (0.7)	5,291	83.6 (0.6)	<0.1 (0.9)	0.96

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. The weighted response rates account for initial sample design as well as CAPI subsampling.

When analyzing the unit response rates by mode of data collection, the only modal comparison that shows a statistically significant difference is the mail response rate. The control treatment had a higher mail response (19.6 percent) than the test treatment (18.4 percent) by 1.2 percentage points. As a result of this difference, we looked at how mail responses differed in the high and low response areas. Table 6 shows the mail response rates for both treatments in high and low response areas.[35] The difference in mail response rates appears to be driven by the difference of rates in the high response areas.

It is possible that the difference in the mail response rates between control and test is related to the content changes made to the test questions. There are some test questions that could be perceived as being too sensitive by some respondents (such as the test question relating to same-sex relationships) and some test questions that could be perceived to be too burdensome by some respondents (such as the new race questions with added race categories). In the automated modes (internet, CATI, and CAPI) there is a higher likelihood of obtaining a sufficient partial response (obtaining enough information to be deemed a response for calculations before the respondent stops answering questions) than in the mail mode. If a respondent is offended by the questionnaire or feels that the questions are too burdensome, they may just throw the questionnaire away and not respond by mail. This could be a possible explanation for the unit response rate being lower for test than control in the mail mode.

[35] Table D-1 (including all modes) can be found in Appendix D.

We note that differences between overall and total self-response response rates were not statistically significant. As most analysis was conducted at this level, we are confident the response rates were sufficient to conduct topic-specific comparisons between the control and test treatments and that there are no underlying response rate concerns that would impact those findings.

Table 6. Mail Response Rates by Designated High (HRA) and Low (LRA) Response Areas

	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
HRA	2,082	20.0 (0.4)	2,224	21.5 (0.4)	-1.5 (0.6)	0.02*
LRA	2,881	13.8 (0.3)	2,948	14.1 (0.3)	-0.3 (0.4)	0.43
Difference		6.2 (0.5)		7.4 (0.4)	-1.1 (0.7)	0.11

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. The weighted response rates account for initial sample design as well as CAPI subsampling.

5.1.2 Unit Response Rates for the Content Follow-Up Interview

Table 7 shows the unit response rates for the CFU interview by mode of data collection of the original interview and for all modes combined, for control and test treatments. Overall, the differences in CFU response rates between the treatments are not statistically significant. The rate at which CAPI respondents from the original interview responded to the CFU interview is lower for test (34.8 percent) than for control (37.7 percent) by 2.9 percentage points. While the protocols for conducting CAPI and CFU were the same between the test and control treatments, we could not account for personal interactions that occur in these modes between the respondent and interviewer. This can influence response rates. We do not believe that the difference suggests any underlying CFU response issues that would negatively affect topic-specific response reliability analysis for comparing the two treatments.

Table 7. Content Follow-Up Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode of Original Interview

Original Interview Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
All Modes	7,867	44.8 (0.5)	7,903	45.7 (0.6)	-0.8 (0.8)	0.30
Internet	4,078	51.9 (0.6)	4,045	52.5 (0.7)	-0.6 (0.8)	0.49
Mail	2,202	46.4 (0.9)	2,197	44.2 (0.9)	2.1 (1.3)	0.11
CATI	369	48.9 (1.9)	399	51.5 (2.5)	-2.5 (2.9)	0.39
CAPI	1,218	34.8 (1.2)	1,262	37.7 (1.1)	-2.9 (1.6)	0.07*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

5.1.3 Demographic and Socioeconomic Profile of Responding Households

One of the underlying assumptions of our analyses in this report is that the sample for the Content Test was selected in such a way that responses from both treatments would be

comparable. We did not expect the demographics of the responding households for control and test treatments to differ. To test this assumption, we calculated distributions for respondent data for the following response categories: *age, sex, educational attainment,* and *tenure.*[36] The response distribution calculations can be found in Table 8. Items with missing data were not included in the calculations. After adjusting for multiple comparisons, none of the differences in the categorical response distributions shown below are statistically significant.

Table 8. Response Distributions: Test versus Control Treatment

Item	Test Percent	Control Percent	Adjusted P-Value
AGE	(n=43,236)	(n=43,325)	0.34
Under 5 years old	5.7 (0.2)	6.1 (0.2)	
5 to 17 years old	17.8 (0.3)	17.6 (0.3)	
18 to 24 years old	8.6 (0.3)	8.1 (0.3)	
25 to 44 years old	25.1 (0.3)	26.2 (0.3)	
45 to 64 years old	26.8 (0.4)	26.6 (0.4)	
65 years old or older	16.0 (0.3)	15.4 (0.3)	
SEX	(n=43,374)	(n=43,456)	1.00
Male	48.8 (0.3)	49.1 (0.3)	
Female	51.2 (0.3)	50.9 (0.3)	
EDUCATIONAL ATTAINMENT#	(n=27,482)	(n=27,801)	1.00
No schooling completed	1.3 (0.1)	1.2 (0.1)	
Nursery to 11th grade	8.1 (0.3)	8.0 (0.3)	
12th grade (no diploma)	1.7 (0.1)	1.6 (0.1)	
High school diploma	21.7 (0.4)	22.3 (0.4)	
GED† or alternative credential	3.5 (0.2)	3.6 (0.2)	
Some college	21.0 (0.4)	20.2 (0.4)	
Associate's degree	8.8 (0.3)	9.1 (0.3)	
Bachelor's degree	20.9 (0.4)	20.3 (0.4)	
Advanced degree	13.1 (0.3)	13.7 (0.3)	
TENURE	(n=17,190)	(n=17,236)	1.00
Owned with a mortgage	43.1 (0.6)	43.2 (0.5)	
Owned free and clear	21.1 (0.4)	21.2 (0.4)	
Rented	33.8 (0.6)	34.0 (0.5)	
Occupied without payment of rent	1.9 (0.2)	1.7 (0.1)	

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
#For ages 25 and older
†General Educational Development
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding.
Significance testing done at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

We also analyzed two other demographic characteristics shown by the responses from the survey: *average household size* and *language of response.* The results for the remaining demographic analyses can be found in Tables 9 and Table 10 below.

[36] We were not able to conduct demographic analysis by relationship status, race, or ethnicity because these topics were tested as part of the Content Test.

Table 9. Comparison of Average Household Size

	Test (n=17,608)	Control (n=17,694)	Test minus Control	P-value
Average Household Size (Number of People)	2.51 (<0.1)	2.52 (<0.1)	>-0.01 (<0.1)	0.76

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a two-tailed t-test at the α=0.1 level.

Table 10. Comparison of Language of Response

Language of Response	Test (n=17,608)	Control (n=17,694)	Test minus Control	P-value
English	96.1 (0.2)	96.2 (0.2)	<0.1 (0.3)	0.52
Spanish	2.7 (0.2)	2.6 (0.2)	<0.1 (0.2)	0.39
Undetermined	1.2 (0.1)	1.2 (0.1)	<0.1 (0.2)	0.62

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a two-tailed t-test at the α=0.1 level.

The Content Test was available in two languages, English and Spanish, for all response options except the mail mode. However, the language of response variable was missing for some responses, so we created a category called "undetermined" to account for those cases.

There are no detectable differences between control and test for *average household size* or *language of response*. There are also no detectable differences for any of the response distributions that we calculated. As a result of these analyses, it appears that respondents in both treatments do exhibit comparable demographic characteristics since none of the resulting findings is significant, which verifies our assumption of demographic similarity between treatments.

5.2 Item Missing Data Rates

*Are the control and test item missing data rates for **Industry** the same across data collection mode and within data collection mode?*

In Table 11, we present the item missing data rates for Industry overall and by mode for the test and control treatments. As noted in our decision criteria, we preferred the Industry item missing data rates in the test treatment to be the same or lower than the control treatment. Overall and for all modes except internet, there was no significant difference between the item missing data rates for the test and control treatments. In the internet mode, the test treatment's rate was 0.6 percentage points higher.

Table 11. Industry Item Missing Data Rates by Mode

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Overall	22,712	5.4 (0.2)	22,973	5.2 (0.2)	0.2 (0.3)	0.58
Internet	11,950	5.3 (0.3)	11,860	4.6 (0.2)	0.6 (0.4)	0.08*
Mail	4,811	10.6 (0.6)	5,126	10.3 (0.6)	0.3 (0.8)	0.67
CATI	889	1.4 (0.5)	869	2.1 (0.6)	-0.7 (0.9)	0.44
CAPI	5062	3.3 (0.4)	5518	3.6 (0.5)	-0.2 (0.6)	0.71

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding.
P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

*Are the control and test item missing data rates for **Occupation** the same across data collection mode and within data collection mode?*

The Occupation item missing data rates are presented in Table 12. Our preference was for the item missing data rates in the test treatment to be the same or lower than the control treatment. The results show there were no significant differences between treatments overall or in any of the modes.

Table 12. Occupation Item Missing Data Rates by Mode

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Overall	22,712	5.4 (0.3)	22,973	5.2 (0.2)	0.2 (0.4)	0.51
Internet	11,950	5.5 (0.3)	11,860	4.9 (0.3)	0.6 (0.4)	>0.10†
Mail	4,811	9.7 (0.6)	5,126	9.7 (0.5)	>-0.1 (0.8)	0.99
CATI	889	1.7 (0.5)	869	2.6 (0.8)	-0.9 (1.0)	0.36
CAPI	5062	3.6 (0.5)	5518	3.5 (0.5)	0.1 (0.7)	0.92

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance was tested based on a two-tailed t-test at the α=0.1 level.
†The actual p-value is greater than 0.10, but rounded to 0.10.

5.3 Codeable Rates

*Are the control and test codeable data rates for **Industry** the same across data collection mode and within data collection mode?*

Table 13 presents the codeable rates for Industry. The codeable rate is the proportion of Industry or Occupation person-level records in universe that were assigned a valid code. Industry and Occupation responses considered to be uncodeable were assigned the code of "9990."[37] We expected the codeable rates for the test treatment to be the same or higher than that of the control treatment. The only significant result was in the mail mode where the codeable data rate for the

[37] Some cases contain Industry or Occupation entries that were considered uncodeable. This indicates that additional research by a referral coder will produce no valid code. Entries such as 'Classified,' 'Don't Know,' 'Blank,' or 'None' could result in an uncodeable entry if not enough information is present in other fields.

test treatment was 0.5 percentage points lower than that of the control treatment. However, the Industry codeable data rate for both treatments exceeds 98.0 percent in all modes.

Table 13. Industry Codeable Data Rates by Mode

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test Minus Control	P-Value
Overall	21,446	99.1 (0.1)	21,694	99.2 (0.1)	-0.1 (0.1)	0.25
Internet	11,362	99.2 (0.1)	11,279	99.3 (0.1)	-0.1 (0.2)	0.54
Mail	4,275	98.3 (0.3)	4,602	98.8 (0.2)	-0.5 (0.3)	0.07*
CATI	875	99.4 (0.4)	856	99.7 (0.3)	-0.3 (0.4)	0.52
CAPI	4,934	99.2 (0.2)	4,957	99.2 (0.2)	>-0.1 (0.3)	0.87

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding.
P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

*Are the control and test codeable data rates for **Occupation** the same across data collection mode and within data collection mode?*

Table 14 displays the codeable data rates for Occupation. We expected the codeable data rates for the test treatment to be the same or higher than that of the control treatment. However, the test treatment had significantly lower codeable rates overall (0.4 percentage points lower) and in the mail mode (0.8 percentage points lower).[38] But the codeable data rates for both treatments are 97.7 percent or higher in all modes.

Table 14. Occupation Item Codeable Data Rates by Mode

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Overall	21,446	98.3 (0.1)	21,694	98.7 (0.1)	-0.4 (0.2)	0.02*
Internet	11,362	98.4 (0.1)	11,279	98.7 (0.1)	-0.3 (0.2)	0.22
Mail	4,275	97.7 (0.3)	4,602	98.5 (0.2)	-0.8 (0.4)	0.06*
CATI	875	98.7 (0.4)	856	99.2 (0.4)	-0.5 (0.6)	0.34
CAPI	4,934	98.4 (0.3)	4,957	98.8 (0.2)	-0.4 (0.4)	0.25

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding.
P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

5.4 Interim Referral Rates

*Are the control and test interim referral rates for **Industry/Occupation** the same across data collection mode and within data collection mode?*

As previously mentioned, the process of coding Industry and Occupation responses occurs simultaneously. When a clerical coder was not able to assign a code to Industry, Occupation, or both, the case was referred to a Referralist (expert coder) where both industry and occupation codes were then subject to change. Referral coders have access to additional resource materials,

[38] Subsequent analysis of test treatment cases receiving the code '9990' in mail mode revealed the majority of respondents did not answer either of the Occupation write-in fields as opposed to not providing enough information to assign an occupation code.

including the NAICS and SOC Manuals as well as internet websites to help determine the most accurate code. We expected the test treatment to have the same or lower referral rates compared to the control treatment.

Table 15 presents the interim referral rates. Overall and for all modes, except internet, the test treatment had a significantly higher interim referral rate than the control treatment. These results were contrary to our expectations. One possible explanation for the higher interim referral rates is the increase in the number of characters allowed from 60 to 100 in the Job Duties write-in field in the test version. The increased detailed information might have resulted in the need for additional research by a Referralist to determine the most accurate code.

Table 15. Industry and Occupation Interim Referral Rates

Mode	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Overall	19,105	21.6 (0.4)	19,359	19.7 (0.4)	2.0 (0.6)	<0.01*
Internet	11,362	22.3 (0.5)	11,279	22.2 (0.5)	0.1 (0.7)	0.89
Mail	4,275	23.4 (0.8)	4,602	19.1 (0.7)	4.3 (1.0)	<0.01*
CATI	875	22.6 (2.0)	856	17.0 (1.8)	5.6 (2.7)	0.04*
CAPI	2,593	18.2 (1.1)	2,622	14.3 (1.0)	4.0 (1.5)	0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding.
P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level.

5.5 Response Distributions

Are the control and test distributions of eligible persons among the NAICS Industry sectors (as indicated by the industry code) the same across data collection mode and within data collection mode?

We compared the Industry response distributions among major NAICS Industry sectors using a Rao-Scott chi-squared test. We expected no difference between the control and test treatments since an increase in details for the industry write-in questions would not necessarily cause a NAICS Industry sector to change. Our expectation of no difference in the NAICS distributions between the control and test treatments is confirmed in the overall results in Table 16 (p-value =0.15). Tables D-5 through D-7 in Appendix D present the results by mode. In the internet mode, the response distributions differed in the following categories: *Wholesale Trade* and *Other Public Services, except Public Administration*.

Table 16. Industry – Response Distribution for Overall

Category	Test Percent (n=21,139)	Control Percent (n=21,459)
Agriculture, forestry, fishing and hunting	1.2 (0.1)	1.3 (0.1)
Mining quarrying, and oil and gas extraction	0.5 (0.1)	0.4 (<0.1)
Construction	6.3 (0.3)	6.2 (0.2)
Manufacturing	10.1 (0.3)	9.4 (0.3)
Wholesale Trade	2.4 (0.2)	2.9 (0.2)
Retail Trade	11.9 (0.3)	11.1 (0.3)
Transportation and Warehousing	3.7 (0.2)	4.2 (0.2)
Utilities	0.8 (0.1)	0.8 (0.1)
Information	2.0 (0.1)	2.0 (0.1)
Finance and insurance	4.4 (0.2)	4.7 (0.2)
Real estate and rental and leasing	2.2 (0.1)	1.9 (0.1)
Professional, scientific, and technical services	7.5 (0.2)	7.4 (0.3)
Management of companies and enterprises	0.1 (<0.1)	0.1 (<0.1)
Administrative and support and waste management services	4.2 (0.2)	4.2 (0.2)
Educational services	9.5 (0.3)	9.1 (0.3)
Health care and social assistance	13.5 (0.3)	13.9 (0.4)
Arts, entertainment and recreation	2.4 (0.2)	2.4 (0.2)
Accommodation and food services	7.6 (0.3)	7.2 (0.3)
Other public services, except public administration	4.7 (0.2)	5.2 (0.3)
Public administration	4.5 (0.2)	5.0 (0.2)
Military	0.7 (0.1)	0.6 (0.1)
Total	100.0	100.0

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: $\chi 2$ =26.6, p-value=0.15Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance testing was conducted at the α=0.1 level.

*Are the control and test distributions of eligible persons among the **SOC** major groups (as indicated by the occupation code) the same across data collection mode and within data collection mode?*

For Occupation, we compared the response distributions among the major SOC groups. We expected no difference between the control and test treatments because an increase in occupational specificity would not necessarily cause an occupation to switch SOC major groups.

Table 17 presents the results of this comparison, overall (across modes). Tables D-2, D-3, and D-4 in Appendix D present the results by mode — internet, mail, and CAI (CATI/CAPI combined to ensure sufficiently large cell sizes), respectively. The test version had no significant effect on the response distribution overall (p-value = 0.43) or in any of the other modes.

Table 17. Occupation – Response Distribution for Overall

Category	Test Percent (n=20,979)	Control Percent (n=21,341)
Management occupations	11.4 (0.3)	10.5 (0.3)
Business and financial operations occupations	5.0 (0.2)	5.3 (0.2)
Computer and mathematical occupations	2.9 (0.2)	3.2 (0.2)
Architecture and engineering occupations	2.1 (0.1)	1.9 (0.1)
Life, physical, and social services occupations	1.1 (0.1)	1.0 (0.1)
Community and social services occupations	1.8 (0.1)	1.7 (0.1)
Legal occupations	1.3 (0.1)	1.1 (0.1)
Education, training, and library occupations	6.3 (0.2)	6.3 (0.2)
Arts, design, entertainment, sports, and media occupations	2.3 (0.2)	2.1 (0.1)
Healthcare practitioner and technical occupations	6.0 (0.2)	6.2 (0.3)
Healthcare support occupations	2.4 (0.2)	2.1 (0.2)
Protective service occupations	2.1 (0.1)	2.2 (0.2)
Food preparation and serving related occupations	5.6 (0.3)	5.7 (0.2)
Building and grounds cleaning and maintenance occupations	3.4 (0.2)	3.8 (0.2)
Personal care and service occupations	3.5 (0.2)	3.6 (0.2)
Sales and related occupations	10.3 (0.3)	10.0 (0.3)
Office and administrative support occupations	12.1 (0.3)	13.2 (0.3)
Farming, fishing and forestry occupations	0.7 (0.1)	0.6 (0.1)
Construction and extraction occupations	4.6 (0.2)	4.7 (0.2)
Installation, maintenance, and repair occupations	3.2 (0.2)	2.8 (0.1)
Production occupations	5.5 (0.2)	5.5 (0.2)
Transportation and material moving occupations	5.9 (0.2)	6.0 (0.3)
Military specific occupations	0.3 (0.1)	0.3 (0.1)
Total	100.0	100.0

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: $\chi 2 = 22.5$, p-value=0.43
Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance testing was done at the α=0.1 level.

5.6 Response Error

*Is response reliability for **Industry** better in test than control?*

For housing units that responded to the original Content Test interview and for which we had a telephone number, a CFU telephone reinterview was conducted to measure response error via the GDR and IOI metrics. Table 18 provides the GDRs for each of the NAICS sectors for the test and control treatments. We expected the test version to have a lower response error than the control version for all response categories (i.e., the same or lower GDR values). The test version had a significantly lower response error for the *Other public services, except public administration* sector (0.9 percentage points lower).

Table 18. Industry – Gross Difference Rates (GDR)

Category	Test GDR Percent (n=8,555)	Control GDR Percent (n=8,941)	Test minus Control	Adjusted P-Value
Agriculture, forestry, fishing and hunting	0.6 (0.1)	0.5 (0.1)	<0.1 (0.2)	1.00
Mining quarrying, and oil and gas extraction	0.6 (0.1)	0.3 (0.1)	0.3 (0.2)	1.00
Construction	2.2 (0.3)	2.1 (0.3)	0.2 (0.4)	1.00
Manufacturing	4.0 (0.3)	3.7 (0.3)	0.3 (0.4)	1.00
Wholesale Trade	3.0 (0.3)	2.8 (0.2)	0.2 (0.5)	1.00
Retail Trade	4.1 (0.4)	4.3 (0.3)	-0.1 (0.5)	1.00
Transportation and Warehousing	1.3 (0.2)	1.6 (0.2)	-0.3 (0.3)	1.00
Utilities	0.3 (0.1)	0.4 (0.1)	-0.1 (0.1)	1.00
Information	1.2 (0.2)	1.1 (0.1)	0.1 (0.2)	1.00
Finance and insurance	0.9 (0.2)	1.0 (0.2)	-0.1 (0.2)	1.00
Real estate and rental and leasing	0.9 (0.1)	1.0 (0.2)	-0.1 (0.2)	1.00
Professional, scientific, and technical services	3.7 (0.3)	3.6 (0.3)	0.1 (0.4)	1.00
Management of companies and enterprises	-	-		-
Administrative and support and waste management services	2.9 (0.3)	2.3 (0.2)	0.5 (0.4)	1.00
Educational services	1.5 (0.2)	1.8 (0.2)	-0.3 (0.2)	1.00
Health care and social assistance	3.0 (0.3)	2.5 (0.2)	0.4 (0.3)	1.00
Arts, entertainment and recreation	1.2 (0.2)	1.1 (0.2)	<0.1 (0.3)	1.00
Accommodation and food services	1.4 (0.2)	1.7 (0.2)	-0.3 (0.3)	1.00
Other public services, except public administration	2.2 (0.2)	3.1 (0.3)	-0.9 (0.3)	0.06*
Public administration	1.6 (0.2)	1.5 (0.2)	0.1 (0.3)	1.00
Military	0.1 (<0.1)	0.4 (0.1)	-0.2 (0.1)	0.15

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test < control) at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. The '-' entry in a cell indicates that either no sample observations or too few sample observations were available to compute an estimate or standard error.

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.

Table 19 presents the IOI for each NAICS Industry sector for the test and control treatments. The *Military* sector in the test treatment had significantly lower response variance than the control treatment (27.8 percentage points lower). The lower response variance was likely due to moving the Armed Forces checkbox from the Employer Name to the Class of Worker question.[39] In addition, specifically asking for the name of the branch of the Armed Forces in the Employer Name question probably increased question clarity.

[39] During the 2016 ACS Content Test, the Class of Worker and the Industry and Occupation questions were tested concurrently.

Table 19. Industry – Index of Inconsistency (IOI)

Category	Test IOI Percent (n=8,555)	Control IOI Percent (n=8,941)	Test minus Control	Adjusted P-Value
Agriculture, forestry, fishing and hunting	22.5 (5.0)	22.5 (5.0)	<0.1 (6.4)	1.00
Mining quarrying, and oil and gas extraction	48.9 (8.9)	31.9 (8.4)	16.9 (12.1)	1.00
Construction	18.3 (2.1)	18.7 (2.3)	-0.4 (3.2)	1.00
Manufacturing	23.1 (1.7)	20.4 (1.6)	2.6 (2.2)	1.00
Wholesale Trade	55.2 (4.6)	48.7 (3.8)	6.5 (6.1)	1.00
Retail Trade	19.8 (1.6)	22.4 (1.4)	-2.6 (2.2)	1.00
Transportation and Warehousing	18.6 (2.8)	21.8 (2.8)	-3.2 (3.8)	1.00
Utilities	19.1 (5.1)	22.8 (5.4)	-3.7 (7.3)	1.00
Information	29.1 (3.8)	25.4 (2.7)	3.6 (4.5)	1.00
Finance and insurance	10.8 (1.8)	10.9 (1.7)	-0.1 (2.4)	1.00
Real estate and rental and leasing	21.8 (3.3)	28.5 (5.5)	-6.6 (6.5)	1.00
Professional, scientific, and technical services	23.6 (1.8)	24.7 (2.2)	-1.1 (2.6)	1.00
Management of companies and enterprises	-	-		-
Administrative and support and waste management services	39.8 (3.4)	31.3 (2.9)	8.5 (4.4)	1.00
Educational services	8.3 (0.9)	9.5 (0.9)	-1.2 (1.3)	1.00
Health care and social assistance	12.8 (1.1)	10.7 (0.7)	2.2 (1.3)	1.00
Arts, entertainment and recreation	25.0 (3.5)	24.3 (3.7)	0.7 (5.5)	1.00
Accommodation and food services	11.2 (1.8)	14.5 (2.0)	-3.2 (2.6)	1.00
Other public services, except public administration	24.9 (2.8)	32.8 (2.6)	-7.9 (3.7)	0.35
Public administration	19.5 (2.6)	15.6 (1.7)	3.9 (3.1)	1.00
Military	8.3 (3.0)	36.2 (9.6)	-27.8 (10.0)	0.06*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test < control) at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. An '-' entry in a cell indicates that either no sample observations or too few sample observations were available to compute an estimate or standard error.

The overall Industry L-fold IOI is reported in Table 20. There were no significant differences between the L-fold IOI values for the test and control treatments.

Table 20. Industry – L-fold Index of Inconsistency (IOI)

Category	Test IOI Percent (n=8,426)	Control IOI Percent (n=8,861)	Test minus Control	P-Value
L-Fold	19.9 (0.6)	20.1 (0.7)	-0.1 (0.9)	0.45

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a one-tailed t-test (test < control) at the α=0.1 level.

*Is response reliability for **Occupation** better in test than control?*

Table 21 displays the GDRs for each of the major SOC groups for the test and control treatments. We expected the test version to have the same or lower response error than the control treatment for all response categories (i.e., the same or lower GDR values). The test

version had a significantly smaller response error for the *Personal care and service* occupations SOC group (0.7 percentage points lower).

Table 21. Occupation – Gross Difference Rates (GDR)

Category	Test GDR Percent (n=8,426)	Control GDR Percent (n=8,861)	Test minus Control	Adjusted P-Value
Management occupations	7.8 (0.5)	8.7 (0.4)	-0.9 (0.6)	1.00
Business and financial operations occupations	4.2 (0.3)	4.3 (0.3)	-0.1 (0.4)	1.00
Computer and mathematical occupations	1.6 (0.2)	1.9 (0.2)	-0.3 (0.3)	1.00
Architecture and engineering occupations	1.3 (0.2)	1.5 (0.2)	-0.2 (0.3)	1.00
Life, physical, and social services occupations	0.8 (0.2)	1.0 (0.2)	-0.2 (0.2)	1.00
Community and social services occupations	1.3 (0.2)	1.0 (0.1)	0.3 (0.3)	1.00
Legal occupations	0.4 (0.1)	0.3 (0.1)	0.1 (0.1)	1.00
Education, training, and library occupations	1.7 (0.2)	1.8 (0.2)	>-0.1 (0.3)	1.00
Arts, design, entertainment, sports, and media occupations	1.3 (0.2)	1.3 (0.2)	-0.1 (0.3)	1.00
Healthcare practitioner and technical occupations	1.9 (0.2)	2.1 (0.2)	-0.2 (0.3)	1.00
Healthcare support occupations	1.6 (0.3)	1.2 (0.2)	0.3 (0.3)	1.00
Protective service occupations	0.6 (0.1)	0.5 (0.1)	0.1 (0.2)	1.00
Food preparation and serving related occupations	1.6 (0.2)	1.8 (0.2)	-0.2 (0.3)	1.00
Building and grounds cleaning and maintenance occupations	1.3 (0.2)	1.6 (0.2)	-0.2 (0.3)	1.00
Personal care and service occupations	1.3 (0.2)	2.0 (0.2)	-0.7 (0.2)	0.09*
Sales and related occupations	5.8 (0.5)	5.6 (0.4)	0.3 (0.6)	1.00
Office and administrative support occupations	7.1 (0.4)	8.2 (0.4)	-1.1 (0.6)	0.62
Farming, fishing and forestry occupations	0.6 (0.1)	0.7 (0.2)	-0.1 (0.2)	1.00
Construction and extraction occupations	2.1 (0.3)	2.3 (0.2)	-0.2 (0.4)	1.00
Installation, maintenance, and repair occupations	2.0 (0.3)	1.7 (0.2)	0.2 (0.3)	1.00
Production occupations	3.2 (0.3)	3.1 (0.3)	>-0.1 (0.4)	1.00
Transportation and material moving occupations	2.8 (0.3)	2.7 (0.3)	0.1 (0.4)	1.00
Military specific occupations	0.2 (0.1)	0.2 (0.1)	<0.1 (0.1)	1.00

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test < control) at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 22 provides the IOI for each major SOC group for the test and control treatments. The simple response variance was high for *Farming, fishing and forestry* occupations in the control treatment. The remaining categories in both treatments had moderate or low response variances. There were no significant differences between the control and treatment IOI values.

Table 22. Occupation – Index of Inconsistency (IOI)

Category	Test IOI Percent (n=8,426)	Control IOI Percent (n=8,861)	Test minus Control	Adjusted P-Value
Management occupations	36.8 (2.0)	43.7 (1.9)	-6.9 (2.7)	0.11
Business and financial operations occupations	41.7 (2.7)	42.3 (2.7)	-0.6 (3.5)	1.00
Computer and mathematical occupations	25.4 (2.8)	26.2 (3.0)	-0.8 (3.9)	1.00
Architecture and engineering occupations	28.9 (3.6)	36.7 (4.0)	-7.7 (5.3)	1.00
Life, physical, and social services occupations	37.1 (5.3)	48.1 (6.6)	-11.1 (7.5)	1.00
Community and social services occupations	32.6 (5.0)	26.2 (3.7)	6.4 (6.3)	1.00
Legal occupations	13.0 (3.0)	11.9 (2.9)	1.1 (4.5)	1.00
Education, training, and library occupations	14.1 (1.8)	13.4 (1.6)	0.7 (2.1)	1.00
Arts, design, entertainment, sports, and media occupations	24.4 (3.5)	30.3 (3.8)	-5.9 (5.0)	1.00
Healthcare practitioner and technical occupations	16.9 (1.9)	19.0 (1.9)	-2.1 (2.8)	1.00
Healthcare support occupations	32.9 (4.7)	31.5 (4.7)	1.3 (6.5)	1.00
Protective service occupations	12.6 (2.4)	11.2 (2.2)	1.4 (3.5)	1.00
Food preparation and serving related occupations	18.0 (2.6)	17.8 (2.3)	0.2 (3.1)	1.00
Building and grounds cleaning and maintenance occupations	21.9 (3.2)	23.3 (3.0)	1.3 (4.3)	1.00
Personal care and service occupations	24.3 (3.0)	30.9 (3.1)	-6.6 (4.1)	1.00
Sales and related occupations	31.8 (2.3)	32.2 (1.8)	-0.5 (2.7)	1.00
Office and administrative support occupations	33.9 (1.8)	36.6 (1.8)	2.7 (2.5)	1.00
Farming, fishing and forestry occupations	38.1 (6.3)	55.6 (8.7)	-17.6 (12.0)	1.00
Construction and extraction occupations	24.2 (2.6)	27.5 (2.8)	-3.3 (3.9)	1.00
Installation, maintenance, and repair occupations	30.2 (4.0)	29.2 (3.3)	1.0 (5.1)	1.00
Production occupations	32.3 (3.1)	32.0 (2.7)	0.3 (4.2)	1.00
Transportation and material moving occupations	27.8 (2.8)	26.1 (2.3)	1.7 (3.9)	1.00
Military specific occupations	37.5 (14.0)	44.1 (14.6)	-6.6 (23.4)	1.00

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a one-tailed t-test (test < control) at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

The L-fold IOI provides an overall response reliability of a question that has multiple mutually exclusive response categories. The L-fold IOI value for the test version was 1.6 percentage points lower than the control version.

Table 23. Occupation – L-fold Index of Inconsistency (IOI)

Category	Test IOI Percent (n=8,426)	Control IOI Percent (n=8,861)	Test minus Control	P-Value
L-Fold	28.2 (0.8)	29.8 (0.8)	-1.6 (1.1)	0.06*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test < control) at the α=0.1 level.

5.7 Analysis of Industry and Occupation Write-in Fields

*For each of the four **Industry and Occupation** write-in fields, is the mean character count for test greater than for control? (overall and by mode)*

The mean character count is the average count of characters provided by respondents in a write-in Industry or Occupation field. There are two fields for Industry and two for Occupation. Characters include spaces, numbers, question marks, exclamation points, other punctuation marks. With the exception of Employer Name, we expected the mean character of the write-in fields to be higher for test than for control treatment due to an expected increase in number of characters collected in the Job Duties questions, and the changes made to the question text and examples. Tables 24 through 27 display the mean character count results of the four individual write-in fields for Industry and Occupation.

The Employer Name question asks respondents to provide the name of their employer. Overall, and in all modes except CATI, the test treatment mean was not significantly higher than the control treatment (see Table 24). In CATI, the mean character count for the test treatment was significantly higher by 1.5 characters.

Table 24. Employer Name – Mean Character Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	20,205	17.1 (0.1)	20,340	17.3 (0.1)	-0.2 (0.1)	0.94
Internet	11,054	18.3 (0.1)	10,996	18.4 (0.1)	-0.1 (0.2)	0.72
Mail	3,876	16.5 (0.2)	4,119	16.7 (0.2)	-0.2 (0.3)	0.78
CATI	806	18.5 (0.4)	772	17.1 (0.5)	1.5 (0.6)	0.01*
CAPI	4,469	15.7 (0.2)	4,453	16.2 (0.2)	-0.5 (0.2)	0.98

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

The Kind of Business question asks respondents to describe the activity, product, or service provided at the location where they are employed. Overall and for all modes except CAPI, the mean character count was significantly higher in the test treatment than the control treatment (see Table 25).

Table 25. Kind of Business – Mean Character Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	20,801	14.2 (0.1)	21,238	13.9 (0.1)	0.3 (0.1)	0.01*
Internet	11,137	14.0 (0.1)	11,138	13.9 (0.1)	0.2 (0.1)	0.08*
Mail	3,962	13.2 (0.1)	4,391	12.7 (0.1)	0.5 (0.2)	0.01*
CATI	866	16.3 (0.5)	848	15.2 (0.5)	1.0 (0.7)	0.06*
CAPI	4,836	14.7 (0.2)	4,861	14.6 (0.2)	0.2 (0.3)	0.27

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

The Job Title question asked respondents for their "main occupation" in the test treatment and "what kind of work" in the control treatment. As illustrated in Table 26, the mean character count in the Job Title write-in field was significantly higher in the test treatment overall and for all modes except internet.

Table 26. Job Title – Mean Character Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	21,078	14.6 (0.1)	21,412	14.1 (0.1)	0.5 (0.1)	<0.01*
Internet	11,197	14.5 (0.1)	11,158	14.3 (0.1)	0.2 (0.2)	0.12
Mail	4,166	13.4 (0.1)	4,527	12.7 (0.1)	0.7 (0.2)	<0.01*
CATI	858	15.0 (0.5)	841	13.8 (0.4)	1.3 (0.6)	0.01*
CAPI	4,857	15.1 (0.2)	4,886	14.5 (0.2)	0.6 (0.3)	0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

The Job Duties question asked respondents to describe their most important activities or duties. Overall and for all modes, the test treatment had a significantly higher mean character count than the control treatment (see Table 27).

Table 27. Job Duties – Mean Character Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	19,964	31.3 (0.3)	20,834	21.0 (0.1)	10.3 (0.3)	<0.01*
Internet	10,426	36.2 (0.3)	10,849	20.6 (0.2)	15.6 (0.4)	<0.01*
Mail	3,882	28.1 (0.5)	4,333	15.8 (0.2)	12.3 (0.5)	<0.01*
CATI	841	28.4 (1.0)	824	23.1 (0.7)	5.4 (1.3)	<0.01*
CAPI	4,815	27.0 (0.4)	4,828	23.7 (0.3)	3.4 (0.5)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

*For each of the four **Industry and Occupation** write-in fields, is the mean word count for test greater than for control? (overall and by mode)*

Since the mean character counts include spaces, numbers, question marks, exclamation points, other punctuation marks, a simple increase in the mean character count itself does not automatically imply a greater detailed response. We also performed a mean word count to determine whether the increases in the mean character were due to punctuation marks or an actual word increase. Table 28 through 31 display the mean word count results of the four individual write-in fields.

For the Employer Name write-in field, overall and for all modes except CATI, the mean word count was not significantly higher for the test than the control treatment.

Table 28. Employer Name – Mean Word Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	20,205	2.6 (0.1)	20,340	2.6 (<0.1)	>-0.1 (<0.1)	0.91
Internet	11,054	2.8 (<0.1)	10,996	2.8 (<0.1)	>-0.1 (<0.1)	0.84
Mail	3,876	2.6 (<0.1)	4,119	2.6 (<0.1)	>-0.1 (<0.1)	0.64
CATI	806	2.8 (0.1)	772	2.6 (0.1)	0.2 (0.1)	0.02*
CAPI	4,469	2.4 (<0.1)	4,453	2.5 (<0.1)	>-0.1 (<0.1)	0.92

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

Table 29 show that overall and for all modes except CAPI, the test treatment had a significantly higher mean word count for Kind of Business. The increase may be the result of the modifications to improve the clarity of the examples. If the increase in mean word count corresponds to an increase in the level of useful detail provided (i.e., more specificity), then it may help clerical coders determine and assign a more accurate industry code.

Table 29. Kind of Business – Mean Word Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	20,801	1.9 (<0.1)	21,238	1.9 (<0.1)	<0.1 (<0.1)	0.01*
Internet	11,137	1.8 (<0.1)	11,138	1.8 (<0.1)	<0.1 (<0.1)	0.04*
Mail	3,962	1.8 (<0.1)	4,391	1.7 (<0.1)	0.1 (<0.1)	<0.01*
CATI	866	2.3 (0.1)	848	2.1 (0.1)	0.2 (0.1)	0.03*
CAPI	4,836	2.0 (<0.1)	4,861	2.0 (<0.1)	<0.1 (<0.1)	0.32

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

Table 30 contains the mean word count for the Job Title write-in for each treatment. Overall, and for all modes, the mean word count was significantly higher for the test treatment than the control treatment. Along with the changes to the question text, the examples listed under each question were also modified in the test treatment. As with Industry, if the increase in mean word count corresponds to an increase in the level of useful detail (i.e., more specificity), then respondents are including details about their Job Title that may help clerical coders in coding occupations.

Table 30. Job Title – Mean Word Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test Minus Control	P-Value
Overall	21,078	2.0 (<0.1)	21,412	1.9 (<0.1)	0.1 (<0.1)	<0.01*
Internet	11,197	1.9 (<0.1)	11,158	1.9 (<0.1)	<0.1 (<0.1)	0.03*
Mail	4,166	1.9 (<0.1)	4,527	1.7 (<0.1)	0.2 (<0.1)	<0.01*
CATI	858	2.1 (0.1)	841	1.9 (0.1)	0.2 (<0.1)	0.02*
CAPI	4,857	2.1 (<0.1)	4,886	2.0 (<0.1)	0.1 (<0.1)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

Responses to the Job Duties write-in field are of particular interest since the test version expanded the number of allowable characters from 60 to 100 in all modes. Respondents were visually cued on internet and mail modes by the additional lines provided for responses. This visual cue was not available to the respondents in the interviewer modes of CATI and CAPI. However, the field representatives conducting the interviews were able enter up to 100 characters in test treatment's instrument field. Overall, and for all modes, the mean word count was significantly higher for the test treatment than the control. As expected, the expansion in the number of allowable characters for the test treatment resulted in significantly higher mean word count when respondents described their main duties.

Table 31. Job Duties – Mean Word Count by Mode

Mode	Test Sample Size	Test Mean	Control Sample Size	Control Mean	Test minus Control	P-Value
Overall	19,964	4.5 (<0.1)	20,834	2.9 (<0.1)	1.6 (<0.1)	<0.01*
Internet	10,426	5.2 (<0.1)	10,849	2.8 (<0.1)	2.5 (0.1)	<0.01*
Mail	3,882	4.1 (0.1)	4,333	2.2 (<0.1)	1.9 (0.1)	<0.01*
CATI	841	4.3 (0.2)	824	3.4 (0.1)	0.9 (0.2)	<0.01*
CAPI	4,815	3.9 (0.1)	4,828	3.4 (<0.1)	0.5 (0.1)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses.
P-values with an asterisk (*) indicate a significant difference based on a one-tailed t-test (test>control) at the α=0.1 level.

*Do the changes to the **Industry and Occupation** questions result in more specificity in the four write-in responses by treatment, across data collection mode, and within data collection mode?*

One goal of the 2016 ACS Content Test was to improve occupational specificity by obtaining more detailed write-in responses. Specificity can be improved if the changes to the question make the coding process easier, more consistent, and improve accuracy of classification. The hope is that the changes will help the Census Bureau provide more detailed coding for industry and occupation than is now available using the current SOC and NAICS aggregations. To address this research question, we compared the level of detail and quality of the write-in responses to the Industry and Occupation questions in both treatments within selected occupation categories by mode of collection. This analysis was conducted for each of the four write-in fields individually, for the two Industry write-in fields combined, and for the two Occupation write-in fields combined. Referralists at NPC conducted the expert qualitative coding review. This review

included cases in 39 different occupation code categories and *Active Duty Military* and *National Guard*, with a total of 6,974 cases. Table C-1 in Appendix C shows the full list of occupations included in the review. Table C-2 shows the distribution of cases reviewed by mode of data collection.

The following is a summary of what Referralists found from the eight questions used as a guide (see Section 2.5.7.3 – Specificity). For Job Duties, most Referralists found that the test treatment write-ins contained the same level of detail or more compared to the control treatment. In particular, for *Management* occupations, the write-in responses for the test treatment often included specifics on the kinds of management, which are a critical element to determine the most accurate occupation code. In addition, a Referralist stated that the details in the write-in responses for *Management* occupations in the test treatment also provided information that assisted in accurately coding Industry responses. One example mentioned by the Referralist was that the details used to classify a write-in as a *Farmers, ranchers, and other agricultural managers*, also helped determine the correct industry code. For *Construction managers,* the industry descriptions in Kind of Business in the test treatment were also more specific.

Another Referralist mentioned the write-in responses for *Teacher* occupations were also more detailed in the test treatment than the control. The Referralist stated the critical elements for coding *Teacher* occupations are details such as school name, grade level, and *Teacher* category. As a result, the great amount of detail in the test treatment would result in a more precise occupation code. In addition, the Kind of Business description and type for *Teachers*, helped in the coding of industry in test treatment versus control treatment. However, the same Referralist stated the responses for *Software developers, applications and systems software and web developers* occupations were more detailed in the control treatment than the test treatment.

Many of the responses to the scaled questions indicated that there were no specific pieces of information missing from the test version that would have helped to code industry or occupation. Many also showed that the test treatment provided at least the same level of detail or more for all four write-in fields.

The expert qualitative coding review indicated that write-in responses in the test treatment had more useful detail compared to the control treatment. Combined with the increase in the mean word count for Industry and Occupation write-in responses in the test treatment, this suggests the question changes and the expansion of the Job Duties write-in field in the test version will improve the quality of the data. The increased specificity of the write-in responses may also lead to the expansion of the number detailed industries and occupations that are published in the Census Bureau industry and occupation code lists and data products.

5.8 Median Coding Time

*Is the median coding time for **Industry/Occupation** for test less than control?*

We compared coding batch times for this analysis. When we explored the distributions of the batch times for both treatments, we found that the data were highly skewed. This occurred due to the way the batch times were calculated. The batch time was calculated from the time the batch was opened until it closed. If the batch was not completed in one day, it was completed the next

day, or in some cases, three days later, especially if a batch was started on a Friday and the coder finished it on Monday. This led to longer batch times for some batches. The quantiles for both distributions are shown in Table D-8 in Appendix D.

Because the data were highly skewed, we compared coding times per batch (in minutes) using the nonparametric Wilcoxon Rank Sum Test. The test results, presented in Table 32, concluded that the coding time (per batch) for the test was not less than that of the control (p-value=0.63) as we had hoped.

Table 32. Median Coding Batch Time (minutes)

Mode	Test Sample Size	Test Median	Control Sample Size	Control Median	P-Value
Overall	124	52.7	126	52.5	0.63

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significance was tested based on a one-tailed t-test (test < control) at the α=0.1 level.

5.9 Industry and Class of Worker Consistency Check

*For the test treatment, how consistent are **Class of Worker** responses with write-in responses about **Industry** (Employer Name and Kind of Business) and the final industry code compared to control? In particular, is the reporting of Active Duty in the Class of Worker test version consistent with the Industry write-in responses and industry code? (overall and by mode)*

The Class of Worker question categorizes workers according to the type of ownership of the employing organization. It is important that a worker's Class of Worker category and industry code be consistent. For example, due to certain NAICS restrictions, anyone with industry code in *Public Administration* must have a Class of Worker category of *Government worker*. The list of tested Class of Worker and industry categories is on Appendix C in Table C-3. Table 33 presents the overall comparisons of the consistency rates of select Class of Worker and industry categories in the control and test treatments. Tables D-9 through D-12 in Appendix D present the results by mode.

The consistency rate for *Private households* in the test treatment was significantly lower overall (5.8 percentage points lower) and in the internet mode (19.8 percentage points lower). The expectation was that for someone working in a *Private households* industry, their Class of Worker category should be either *Private for-profit, Self-employed not incorporated, Self-employed incorporated,* or *Unpaid family workers*. The results indicated that respondents in the test treatment more often selected a category of *Private not-for-profit* or one of the three government worker categories (overall and in the internet mode).

Table 33. Class of Worker and Industry Consistency (in Percent): Overall

Category	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Private Households	123	93.2 (2.5)	151	99.0 (0.6)	-5.8 (2.6)	0.03*
Not-for-profit organizations	295	87.9 (3.0)	329	81.7 (2.6)	6.2 (4.0)	0.12
Public Administration	974	96.1 (0.7)	1,073	96.1 (0.7)	<0.1 (1.0)	0.98
Postal Service	96	80.7 (7.7)	103	67.0 (7.3)	13.8 (11.1)	0.22
Active Duty Military	144	96.8 (3.1)	135	99.4 (0.4)	-2.6 (3.2)	0.42
Military Reserves and National Guard	-	-	-	-	-	-

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significant at α=0.1 level.
P-values with an asterisk (*) indicate a significant difference between the two rates. An entry of '-' in a cell indicates that no or too few observations were available to meet statistical standards for reliability.

6 CONCLUSIONS AND RECOMMENDATIONS

In the 2016 ACS Content Test, we tested a proposed revision to the Industry and Occupation questions. The revisions were designed to improve the clarity of the questions, the accuracy of the responses, and the level of detail provided. There were several key changes to these questions. We tested new, complementary examples for the Kind of Business, Job Title, and Job Duties questions to elicit more detail from the respondents. The Class of Worker, Industry, and Occupation items were tested as a single, multi-part question to show that these were all a part of a series with one general instruction format. For the Job Duties question, the test version was designed to allow the respondent to give more detail. We assessed the proposed changes to the Industry and Occupation questions through a variety of metrics. The analysis of metrics yielded a combination of results.

Overall, there were no significant results for Occupation on item missing data rates. For Industry, while the item missing data rates in internet mode were significantly higher for the test treatment, these were not significantly different overall. The codeable data rates for Industry were significantly lower in the mail mode for the test treatment, but not significantly different overall. For Occupation, the overall codeable rates were significantly lower for the test treatment. However, the overall codeable rates for both the control and test treatment each exceeded 98.0 percent. The interim referral rates for the test treatment were significantly higher overall and for all modes, except internet.

We also assessed the reliability of the questions through a follow-up operation. The response variance for the Military Industry was significantly lower for the test treatment by 27.8 percentage points, most likely attributable to moving the Armed Forces checkbox from the Employer Name to the Class of Worker question. The response error for Industry was also significantly lower in the test treatment for Other Public Services, except Public Administration. For Occupation, the response error was significantly lower in the test treatment for Personal Care and Service Occupations.

One of the main reasons for testing a new version of the Industry and Occupation questions was to improve occupational specificity in the write-in responses. Our Expert Qualitative Coding Review found that the write-in responses to the Job Duties field were more detailed in the test treatment. The increase in the number of characters allowed for responses to the Job Duties question in the test treatment resulted in a significant increase in the mean word and character count of responses overall and for all modes. Our expectation is that this increased detail will improve coding accuracy.

The recommendation of the Industry and Occupation Statistics Branch is to adopt the test version of the Industry and Occupation questions. The results of the 2016 ACS Content Test, most notably the increased specificity of the write-ins responses, may make it possible to produce more accurate and detailed codes for industry and occupation categories than is available using the current coding system for industry and occupation.

7 ACKNOWLEDGEMENTS

The 2016 ACS Content Test would not have been possible without the participation and assistance of many individuals from the Census Bureau and other agencies. Their contributions are sincerely appreciated and gratefully acknowledged.

Census Bureau staff in the American Community Survey Office (ACSO), Application Development and Services Division (ADSD), Decennial Information Technology Division (DITD), Decennial Statistical Studies Division (DSSD), Field Division (FLD), National Processing Center (NPC), Population Division (POP), and Social, Economic, and Housing Statistics Division (SEHSD).

Representatives from other agencies in the Federal statistical system serving on the Office of Management and Budget's Interagency Working Group for the ACS and the Topical Subcommittees formed by the Working Group for each topic tested on the 2016 ACS Content Test.

Staff in the Office of Management and Budget's Statistical and Science Policy Office.

The authors would like to thank the following individuals for their contributions to the analysis and review of the *2016 American Community Survey Content Test Evaluation Report: Industry and Occupation* - Michael Risley (DSSD) for verifying the results of the statistical analysis. We are also grateful to Jennifer Ortman (ACSO), Lynda Laughlin (SEHSD), Anthony Tersine (DSSD), Elizabeth Poehler (DSSD), Kelly Holder (SEHSD), Jennifer Cheeseman Day (SEHSD), Julia B Beckhusen (SEHSD), Nicole Scanniello (SEHSD), and Howard Hogan (DEPDIR) for their guidance in preparing this report and for reviewing it.

Additionally, the authors would like to acknowledge Melissa Chiu (Center for Administrative Records Research and Applications (CARRA)), Liana Christin Landivar (Department of Labor (DOL), formerly Census Bureau), and Mary C. Davis (Center for Survey Measurement (CSM)) for the assistance in the development of the questions to be tested. We are also thankful to Padraic Murphy and Sarah K. Heimel (DSSD), for assistance in the development of the decision criteria and research questions for this Content Test topic. Also, Ivonne Pabón Marrero (retired

Census Bureau) and Héctor Martínez (Decennial Census Management Division (DCMD)) for the translation of the questions being tested and their related materials. Lastly, we would like to recognize the contributions of the cognitive interviewing implementation and analysis team, including Martha Stapleton (Westat).

Representatives not mentioned previously from the Census Bureau and other agencies in the Federal statistical system serving on the Office of Management and Budget's Interagency Working Group for the ACS and the Topical Subcommittee for the Industry and Occupation questions tested on the 2016 ACS Content Test include Dameka Reese (ACSO); J. Kent Patterson from the NPC; John Finamore, from the National Center for Science and Engineering Statistics (NCSES) at the National Science Foundation (NSF); and Joseph Donovan and Elvira Sisolak from the Equal Employment Opportunity Commission (EEOC).

We are thankful to the SEHSD ACS Coordination Branch for their guidance and assistance throughout the 2016 ACS Content Test process. Furthermore, we are thankful to Michael Coan and Tammy Worth (DITD) for processing the Industry and Occupation coding files, the industry and occupation coders and Referralists at NPC, and Traci Least and Rick Downs (DSD) who made the coding process for the Content Test a successful one. Finally, the authors gratefully acknowledge the subject matter expertise assistance of Julia Manzella and the other staff members from the Industry and Occupation Statistics Branch (IOSB) from SEHSD.

8 REFERENCES

Biemer, P. (2011). *Latent Class Analysis of Survey Error*. Wiley, New York.

Bureau of Labor Statistics. (2010). *Standard Occupational Classification System*. Retrieved March 14, 2014, from United States Bureau of Labor Statistics Official Web Site: https://www.bls.gov/soc/major_groups.htm.

Bureau of Labor Statistics. (2015). *Occupational Employment Statistics*. Retrieved September 15, 2016, from United States Bureau of Labor Statistics Official Web Site: https://www.bls.gov/oes/2015/may/oes_stru.htm.

Dusch, G. & Meier, F. (2012). *2010 Census Content Reinterview Survey Evaluation Report,* U.S. Census Bureau, June 13, 2012. Retrieved May 17, 2016, from United States Census Bureau Official Web Site: http://www.census.gov/2010census/pdf/2010_Census_Content_Reinterview_Survey_Evaluation_Report.pdf.

Flanagan, P. (1996). *Survey Quality & Response Variance* (Unpublished Internal Document). U.S. Census Bureau. Demographic Statistical Methods Division. Quality Assurance and Evaluation Branch.

Flanagan, P. (2001). *Measurement Errors in Survey Response*. University of Maryland Baltimore County, Baltimore, Maryland.

Holm, S. (1979). "A Simple Sequentially Rejective Multiple Test Procedure," Scandinavian Journal of Statistics, Vol. 6, No. 2: 65-70. Retrieved on January 31, 2017 from https://www.jstor.org/stable/4615733?seq=1#page_scan_tab_contents.

Raglin, D. (2014). *American Community Survey Fiscal Year 2014 Content Review Interviewer Survey Results.* U.S. Census Bureau.

Rao, J. N. K. & Scott, A. J. (1987). "On Simple Adjustments to Chi-Square Tests with Sample Survey Data," The Annal of Statistics, Vol. 15, No. 1, 385-397. Retrieved on January 31, 2017 from http://projecteuclid.org/euclid.aos/1176350273.

Stapleton, M. & Steiger, D. (2015). *Cognitive Testing of the 2016 American Community Survey Content Test Items: Summary Report for Round 1 and Round 2 Interviews.* Westat, Rockville, Maryland, January 2015.

Steiger, D., Robins, C., Trundle, K., Stapleton, M., Sugovic, M., Plotkin, M., Hamilton, D., Nooraddini, I., Matos, R., Matheus, M. Sherehiy, B., Restrepo, J. (2014). *Cognitive Testing of the 2016 American Community Survey Content Test Items: Briefing Report for Round 1 Interviews.* Westat, Rockville, Maryland, August 2014.

Steiger, D., Robins, C., Stapleton, M., Anderson, J., Beauvais-Dennig, S., Folz, J., Leonard, M., Shkodriani, G. (2015a). *Cognitive Testing of the 2016 American Community Survey Content Test Items: Briefing Report for Round 2 Interviews.* Westat, Rockville, Maryland, January 2015.

Steiger, D., Anderson, J., Folz, J., Leonard, M., & Stapleton, M. (2015b). Cognitive Testing of the 2016 American Community Survey Content Test Items: Briefing Report for Round 3 Interviews. Westat, Rockville, Maryland, June 2015.

Tegler, M., Downs, B., Kirk, M., & Ericson, L. (2007). Evaluation Report Covering Industry and Occupation Items. (2006 American Community Survey Content Test Report P.7). U.S. Census Bureau.

U.S. Census Bureau. (2010). *2010 Census Occupation Code List*. Retrieved March 27, 2014, from United States Census Bureau Official Web Site: http://www.census.gov/people/io/methodology/.

U.S. Census Bureau. (2012). *North American Industry Classification System*. Retrieved March 27, 2014, from United States Census Bureau Official Web Site: https://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2012.

U.S. Census Bureau. (2013). *2012 Census Industry Code List*. Retrieved on March 27, 2014, from United States Census Bureau Official Web Site: http://www.census.gov/people/io/methodology/.

U.S. Census Bureau. (2014). *American Community Survey Design and Methodology (January 2014).* Retrieved February 1, 2017 from United States Census Bureau Official Web Site: http://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

U.S. Census Bureau (2016). *2015 Planning Database Tract Data* [Data file]. Retrieved January 31, 2017 from United States Census Bureau Official Web Site: http://www.census.gov/research/data/planning_database/2015/.

U.S. Office of Management and Budget, Executive Office of the President. (2016). *Notice of Standard Occupational Classification Policy Committee Recommendations to OMB and solicitation of comments.* Federal Register, Vol. 81, No. 141: 48309. Retrieved February 6, 2017, from United States Bureau of Labor Statistics Official Web Site: https://www.bls.gov/soc/2018/frn_july_22_2016.pdf.

Appendix A: Internet Versions of the Control and Test Questions

Figure A- 1.
Control Version of the Industry and Occupation Questions

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)*
|_____|*[60 characters]*

Is this business mainly –
☐ manufacturing?
☐ wholesale trade?
☐ retail trade?
☐ other (agriculture, construction, service, government, etc.)?

What kind of work was (name) doing at this job? *(For example: registered nurse, personnel manager, supervisor of order department, secretary, accountant)*
|_____|*[60 characters]*

What were (Name)'s most important activities or duties? *(For example: patient care, directing hiring policies, supervising order clerks, typing and filing, reconciling financial records)*
|_____|*[60 characters]*

Figure A- 2.
Test Version of the Industry and Occupation Questions

c. What kind of business or industry was this? *Include the main activity, product, or service provided at the location where employed. (For example: elementary school, residential construction)*
|_____| *[60 characters]*

d. Was this mainly --
☐ manufacturing?
☐ wholesale trade?
☐ retail trade?
☐ other (agriculture, construction, service, government, etc.)?

e. What was (Name)'s main occupation? *(For example: 4th grade teacher, entry-level plumber)*
|_____|*[60 characters]*

f. Describe (Name)'s most important activities or duties. *(For example: instruct and evaluate students in math and create lesson plans, assemble and install pipe sections and review building plans for work details)*

_____
|_____|
|_____|
|_____|
[100 characters]

Appendix B: CATI/CFU and CAPI Versions of the Control and Test Questions

Figure B-1. Control Version of the Industry and Occupation Questions – CATI and CAPI

What kind of business or industry was this?

For example: hospital, newspaper publishing, mail order house, auto engine manufacturing, bank.

What kind of work < was (Name)/were you> doing at this job?

For example: registered nurse, personnel manager, supervisor of order department, secretary, accountant.

What were <(Name)'s/your> most important activities or duties at this job?
For example: patient care, directing hiring policies, supervising order clerks, typing and filing, reconciling financial records.

Figure B-2. Test Version of the Industry and Occupation Questions – CATI and CAPI

What kind of business or industry was this?
Include the main activity, product, or service provided at the location where employed. For example: elementary school, residential construction, or another kind of business.

Was this mainly – manufacturing, wholesale trade, retail trade, or some other kind of business?

- Manufacturing
- Wholesale trade
- Retail trade
- Other (agriculture, construction, service, government, etc.)

What was <(Name)'s/your> main occupation?

For example: 4th grade teacher, entry-level plumber, or another occupation. Describe <(Name)'s/your> most important activities or duties. For example: instruct and evaluate students and create lesson plans, assemble and install pipe sections and review building plans for work details, or other duties.

Appendix C: Expert Qualitative Coding and Industry and Class of Worker Consistency

Table C-1. List of Occupations – Expert Qualitative Coding

Reviewed Occupation Code	Occupation Descriptions
0010	Chief executives
0020	General and operations managers
0050	Marketing and sales managers
0110	Computer and information systems managers
0120	Financial managers
0136	Human resources managers
0140	Industrial production managers
0150	Purchasing managers
0160	Transportation, storage, and distribution managers
0205	Farmers, ranchers, and other agricultural managers
0220	Construction managers
0230	Education administrators
0300	Architectural and engineering managers
0410	Property, real estate, and community association managers
0430	Managers, all other
1006	Computer systems analysts
1010	Computer programmers
1020	Software developers, applications and systems software
1030	Web developers
1050	Computer support specialists
1060	Database administrators
1105	Network and computer systems administrators
1106	Computer network architects
1107	Computer occupations, all other
2200	Postsecondary teachers
2300	Preschool and kindergarten teachers
2310	Elementary and middle school teachers
2320	Secondary school teachers
2330	Special education teachers
2340	Other teachers and instructors
2540	Teacher assistants
3060	Physicians and surgeons
3600	Nursing, psychiatric, and home health aides
4600	Childcare workers
4610	Personal care aides
5240	Customer service representatives
5700	Secretaries and administrative assistants
9130	Driver/sales workers and truck drivers
5700	Secretaries and administrative assistants
AD	Active Duty Military cases
NG	National Guard cases

Table C-2. Expert Qualitative Coding – Number Of Cases Reviewed, By Mode

Mode	Test Sample Size	Control Sample Size
Overall	3,434	3,540
Internet	1,748	1,796
Mail	874	896
CATI/CAPI	812	848

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Table C-3. Class of Worker and Industry Consistency Check

Class of Worker	**Industry Description**
Private for-profit, Self-employed not incorporated, Self-employed incorporated, or Unpaid family workers	Private Households
Private not-for-profit	Religious organizations, Civic, social, advocacy organizations, and grantmaking and giving services, Labor unions, and Business, professional, political, and similar organizations
State government, Local government, Federal government, or Unspecified government	Public Administration
Federal government	Postal Service
Federal government	U.S. Army, U.S. Air Force, U.S. Navy, U.S. Marines, U.S. Coast Guard, Armed Forces, and Branch not specified
State and Federal government	Military Reserves or National Guard

Appendix D: Additional Tables

Table D-1. Unit Response Rates by Designated High (HRA) and Low (LRA) Response Areas

Mode	Test Interviews	Test Percent	Control Interviews	Control Percent	Test minus Control	P-Value
Total Response	19,400		19,455			
HRA	7,556	94.3 (0.4)	7,608	94.5 (0.3)	-0.2 (0.6)	0.72
LRA	11,844	91.5 (0.3)	11,847	91.0 (0.3)	0.5 (0.5)	0.29
Difference		2.7 (0.5)		3.5 (0.5)	-0.7 (0.7)	0.33
Self-Response	13,131		13,284			
HRA	6,201	59.7 (0.7)	6,272	60.6 (0.7)	-0.9 (0.9)	0.31
LRA	6,930	33.2 (0.4)	7,012	33.6 (0.4)	-0.4 (0.6)	0.55
Difference		26.5 (0.8)		27.0 (0.8)	-0.5 (1.2)	0.66
Internet	8,168		8,112			
HRA	4,119	39.6 (0.6)	4,048	39.1 (0.6)	0.5 (0.8)	0.51
LRA	4,049	19.4 (0.3)	4,064	19.5 (0.3)	0.1 (0.4)	0.87
Difference		20.2 (0.6)		19.6 (0.7)	0.6 (0.9)	0.52
Mail	4,963		5,172			
HRA	2,082	20.0 (0.4)	2,224	21.5 (0.4)	-1.5 (0.6)	0.02*
LRA	2,881	13.8 (0.3)	2,948	14.1 (0.3)	-0.3 (0.4)	0.43
Difference		6.2 (0.5)		7.4 (0.4)	-1.1 (0.7)	0.11
CATI	872		880			
HRA	296	9.0 (0.5)	301	9.6 (0.6)	-0.6 (0.8)	0.44
LRA	576	7.9 (0.4)	579	8.0 (0.3)	-0.1 (0.5)	0.85
Difference		1.1 (0.6)		1.6 (0.7)	-0.5 (0.9)	0.58
CAPI	5,397		5,291			
HRA	1,059	82.2 (1.0)	1,035	82.7 (0.9)	-0.5 (1.3)	0.69
LRA	4,338	85.8 (0.5)	4,256	85.0 (0.4)	0.8 (0.7)	0.23
Difference		-3.7 (1.1)		-2.3 (1.0)	-1.3 (1.5)	0.36

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: Standard errors are in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. The weighted response rates account for initial sample design as well as CAPI subsampling.

Table D-2. Occupation Response Distribution for Internet

Category	Test Percent (n=11,116)	Control Percent (n=11,093)
Management occupations	14.3 (0.4)	13.5 (0.5)
Business and financial operations occupations	6.6 (0.3)	6.7 (0.3)
Computer and mathematical occupations	4.6 (0.3)	4.4 (0.2)
Architecture and engineering occupations	2.9 (0.2)	2.8 (0.2)
Life, physical, and social services occupations	1.5 (0.2)	1.3 (0.1)
Community and social services occupations	2.2 (0.2)	1.8 (0.2)
Legal occupations	1.8 (0.2)	1.4 (0.2)
Education, training, and library occupations	8.2 (0.3)	8.4 (0.3)
Arts, design, entertainment, sports, and media occupations	2.6 (0.2)	2.5 (0.2)
Healthcare practitioner and technical occupations	6.7 (0.3)	6.9 (0.3)
Healthcare support occupations	1.9 (0.1)	1.6 (0.1)
Protective service occupations	2.2 (0.2)	2.0 (0.1)
Food preparation and serving related occupations	3.9 (0.2)	3.9 (0.2)
Building and grounds cleaning and maintenance occupations	2.0 (0.2)	2.0 (0.2)
Personal care and service occupations	2.7 (0.2)	2.8 (0.2)
Sales and related occupations	9.6 (0.3)	10.3 (0.3)
Office and administrative support occupations	12.2 (0.4)	13.1 (0.4)
Farming, fishing and forestry occupations	0.3 (0.1)	0.3 (0.1)
Construction and extraction occupations	3.0 (0.2)	2.9 (0.2)
Installation, maintenance, and repair occupations	2.6 (0.2)	2.3 (0.2)
Production occupations	3.6 (0.2)	4.2 (0.2)
Transportation and material moving occupations	4.1 (0.2)	4.4 (0.2)
Military specific occupations	0.5 (0.1)	0.3 (0.1)
Total	100.0	100.0

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: $\chi^2 = 23.7$, p-value=0.36

Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance testing was done at the α=0.1 level.

Table D-3. Occupation Response Distribution for Mail

Category	Test Percent (n=4,161)	Control Percent (n=4,522)
Management occupations	8.7 (0.6)	8.2 (0.5)
Business and financial operations occupations	4.6 (0.4)	4.6 (0.4)
Computer and mathematical occupations	1.7 (0.3)	2.1 (0.3)
Architecture and engineering occupations	1.8 (0.3)	2.2 (0.3)
Life, physical, and social services occupations	0.9 (0.2)	0.9 (0.2)
Community and social services occupations	1.6 (0.2)	1.8 (0.3)
Legal occupations	1.3 (0.2)	0.9 (0.2)
Education, training, and library occupations	6.5 (0.5)	6.6 (0.5)
Arts, design, entertainment, sports, and media occupations	2.3 (0.3)	2.0 (0.2)
Healthcare practitioner and technical occupations	7.5 (0.6)	6.3 (0.5)
Healthcare support occupations	2.5 (0.3)	2.2 (0.2)
Protective service occupations	2.0 (0.2)	2.1 (0.3)
Food preparation and serving related occupations	4.2 (0.3)	5.5 (0.4)
Building and grounds cleaning and maintenance occupations	2.5 (0.2)	3.9 (0.4)
Personal care and service occupations	3.9 (0.4)	4.1 (0.4)
Sales and related occupations	10.9 (0.6)	10.3 (0.6)
Office and administrative support occupations	14.6 (0.7)	14.5 (0.7)
Farming, fishing and forestry occupations	0.6 (0.2)	0.7 (0.2)
Construction and extraction occupations	5.1 (0.4)	4.5 (0.4)
Installation, maintenance, and repair occupations	3.3 (0.4)	3.6 (0.4)
Production occupations	6.1 (0.5)	6.8 (0.5)
Transportation and material moving occupations	7.1 (0.5)	6.1 (0.4)
Military specific occupations	0.2 (0.1)	0.2 (0.1)
Total	100.0	100.0

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: χ^2= 26.7, p-value=0.22

Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance testing was done at the α=0.1 level.

Table D-4. Occupation Response Distribution for CAI

Category	Test Percent (n=5,702)	Control Percent (n=5,726)
Management occupations	9.2 (0.6)	8.1 (0.6)
Business and financial operations occupations	3.4 (0.4)	4.0 (0.4)
Computer and mathematical occupations	1.3 (0.3)	2.3 (0.4)
Architecture and engineering occupations	1.4 (0.2)	0.7 (0.2)
Life, physical, and social services occupations	0.8 (0.2)	0.8 (0.2)
Community and social services occupations	1.5 (0.3)	1.5 (0.3)
Legal occupations	0.7 (0.2)	0.9 (0.2)
Education, training, and library occupations	4.0 (0.4)	3.9 (0.4)
Arts, design, entertainment, sports, and media occupations	2.0 (0.3)	1.7 (0.3)
Healthcare practitioner and technical occupations	4.4 (0.4)	5.4 (0.6)
Healthcare support occupations	3.0 (0.4)	2.7 (0.3)
Protective service occupations	2.0 (0.3)	2.5 (0.4)
Food preparation and serving related occupations	8.1 (0.6)	7.8 (0.5)
Building and grounds cleaning and maintenance occupations	5.5 (0.5)	5.7 (0.4)
Personal care and service occupations	4.3 (0.4)	4.2 (0.4)
Sales and related occupations	10.8 (0.5)	9.5 (0.6)
Office and administrative support occupations	11.0 (0.6)	12.6 (0.7)
Farming, fishing and forestry occupations	1.0 (0.2)	1.0 (0.2)
Construction and extraction occupations	6.3 (0.5)	6.8 (0.5)
Installation, maintenance, and repair occupations	3.9 (0.4)	3.0 (0.3)
Production occupations	7.4 (0.6)	6.3 (0.5)
Transportation and material moving occupations	7.5 (0.5)	7.8 (0.6)
Military specific occupations	0.2 (0.1)	0.4 (0.2)
Total	100.0	100.0

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: $\chi 2 = 29.9$, p-value=0.12

Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance testing was done at the α=0.1 level.

Table D-5. Industry Response Distribution for Internet

Category	Test Percent (n=11,221)	Control Percent (n=11,156)	Test minus Control	Adjusted P-Value
Agriculture, forestry, fishing and hunting	0.8 (0.1)	0.7 (0.1)	0.1 (0.2)	1.00
Mining quarrying, and oil and gas extraction	0.5 (0.1)	0.4 (0.1)	<0.1 (0.1)	1.00
Construction	5.0 (0.3)	4.1 (0.2)	0.8 (0.4)	0.51
Manufacturing	9.4 (0.3)	10.1 (0.4)	-0.7 (0.5)	1.00
Wholesale Trade	2.2 (0.2)	3.0 (0.2)	-0.8 (0.3)	0.07*****
Retail Trade	10.3 (0.4)	10.1 (0.3)	0.2 (0.5)	1.00
Transportation and Warehousing	3.4 (0.2)	3.6 (0.2)	-0.3 (0.3)	1.00
Utilities	0.9 (0.1)	1.0 (0.1)	-0.1 (0.2)	1.00
Information	2.5 (0.2)	2.7 (0.2)	-0.1 (0.3)	1.00
Finance and insurance	5.9 (0.3)	5.6 (0.3)	0.2 (0.4)	1.00
Real estate and rental and leasing	2.3 (0.2)	2.1 (0.2)	0.3 (0.3)	1.00
Professional, scientific, and technical services	10.6 (0.4)	9.5 (0.3)	1.1 (0.5)	0.68
Management of companies and enterprises	0.1 (<0.1)	0.2 (<0.1)	-0.1 (0.1)	0.58
Administrative and support and waste management services	3.4 (0.2)	3.1 (0.2)	0.4 (0.3)	1.00
Educational services	12.0 (0.4)	11.9 (0.4)	0.1 (0.5)	1.00
Health care and social assistance	13.1 (0.4)	13.3 (0.4)	-0.2 (0.5)	1.00
Arts, entertainment and recreation	2.4 (0.2)	2.5 (0.2)	-0.1 (0.5)	1.00
Accommodation and food services	5.2 (0.2)	5.3 (0.3)	>-0.1 (0.3)	1.00
Other public services, except public administration	3.9 (0.2)	4.9 (0.3)	-1.0 (0.3)	0.07*
Public administration	5.3 (0.3)	5.3 (0.2)	<0.1 (0.4)	1.00
Military	0.7 (0.1)	0.6 (0.1)	0.2 (0.1)	1.00
Total	100.0	100.0		

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: χ^2= 37.7, p-value=0.01

Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. P-values with an asterisk (*) indicate a significant difference based on a two-tailed t-test at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table D-6. Industry Response Distribution for Mail

Category	Test Percent (n=4,161)	Control Percent (n=4,532)
Agriculture, forestry, fishing and hunting	1.2 (0.2)	1.4 (0.3)
Mining quarrying, and oil and gas extraction	0.4 (0.1)	0.3 (0.1)
Construction	6.9 (0.5)	5.7 (0.5)
Manufacturing	11.3 (0.6)	11.0 (0.6)
Wholesale Trade	3.0 (0.4)	2.7 (0.3)
Retail Trade	12.5 (0.8)	12.1 (0.7)
Transportation and Warehousing	4.2 (0.4)	3.8 (0.4)
Utilities	0.6 (0.1)	1.1 (0.2)
Information	1.5 (0.2)	2.1 (0.3)
Finance and insurance	4.4 (0.4)	5.0 (0.4)
Real estate and rental and leasing	2.1 (0.3)	1.9 (0.2)
Professional, scientific, and technical services	5.3 (0.4)	6.1 (0.5)
Management of companies and enterprises	<0.1 (<0.1)	0.1 (0.1)
Administrative and support and waste management services	2.8 (0.3)	3.7 (0.4)
Educational services	9.9 (0.6)	9.1 (0.6)
Health care and social assistance	15.1 (0.7)	14.6 (0.6)
Arts, entertainment and recreation	2.3 (0.3)	2.5 (0.3)
Accommodation and food services	5.8 (0.5)	6.6 (0.5)
Other public services, except public administration	5.4 (0.5)	5.4 (0.5)
Public administration	4.9 (0.4)	4.4 (0.3)
Military	0.3 (0.1)	0.4 (0.1)
Total	100.0	100.0

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: χ^2= 20.1, p-value=0.45

Standard errors are shown in parentheses. Significance testing was done at the α=0.1 level.

Table D-7. Industry Response Distribution for CAI

Category	Test Percent (n=5,757)	Control Percent (n=5,771)	Test minus Control	Adjusted P-Value
Agriculture, forestry, fishing and hunting	1.6 (0.2)	1.8 (0.2)	-0.3 (0.3)	1.00
Mining quarrying, and oil and gas extraction	0.5 (0.1)	0.4 (0.1)	0.1 (0.2)	1.00
Construction	7.6 (0.6)	8.8 (0.5)	-1.1 (0.8)	1.00
Manufacturing	10.4 (0.6)	8.0 (0.6)	2.4 (0.9)	0.12
Wholesale Trade	2.3 (0.3)	2.8 (0.3)	-0.5 (0.4)	1.00
Retail Trade	13.4 (0.6)	11.7 (0.7)	1.7 (1.0)	1.00
Transportation and Warehousing	4.0 (0.4)	5.0 (0.5)	-1.0 (0.6)	1.00
Utilities	0.7 (0.2)	0.4 (0.1)	0.3 (0.2)	1.00
Information	1.7 (0.3)	1.3 (0.2)	0.4 (0.3)	1.00
Finance and insurance	2.6 (0.3)	3.4 (0.3)	-0.9 (0.5)	1.00
Real estate and rental and leasing	2.0 (0.3)	1.7 (0.3)	0.2 (0.4)	1.00
Professional, scientific, and technical services	4.9 (0.4)	5.6 (0.5)	-0.7 (0.7)	1.00
Management of companies and enterprises	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	1.00
Administrative and support and waste management services	5.6 (0.4)	5.9 (0.5)	-0.3 (0.6)	1.00
Educational services	6.4 (0.5)	5.7 (0.5)	0.7 (0.8)	1.00
Health care and social assistance	13.3 (0.8)	14.2 (0.8)	-0.9 (1.1)	1.00
Arts, entertainment and recreation	2.5 (0.3)	2.2 (0.3)	0.3 (0.5)	1.00
Accommodation and food services	11.0 (0.7)	9.8 (0.6)	1.2 (1.0)	1.00
Other public services, except public administration	5.3 (0.4)	5.5 (0.5)	-0.2 (0.7)	1.00
Public administration	3.3 (0.4)	4.9 (0.5)	-1.6 (0.6)	0.19
Military	0.8 (0.2)	0.7 (0.2)	0.1 (0.3)	1.00
Total	100.0	100.0		

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Note: χ^2= 37.4, p-value=0.01

Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance was tested based on a two-tailed t-test at the α=0.1 level. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table D-8. Quantiles for Coding Batch Times (in minutes)

Treatment	Sample Size	First Quantile	Median	Third Quantile
Control	126	34.6	52.6	77.8
Test	124	35.1	52.7	84.6

Source: U.S. Census Bureau, 2016 American Community Survey Content Test

Table D-9. Class of Worker and Industry Consistency (in Percent): Internet

Category	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Private Households	38	79.7 (8.5)	59	99.5 (0.6)	-19.8 (8.5)	0.02*
Not-for-profit organizations	184	91.1 (2.7)	208	85.8 (2.8)	5.3 (3.9)	0.17
Public Administration	590	96.5 (0.9)	601	96.7 (0.8)	-0.2 (1.2)	0.85
Postal Service	45	88.4 (7.4)	52	67.9 (9.2)	20.4 (12.8)	0.11
Active Duty Military	96	100.0 (0.0)	95	99.5 (0.5)	0.5 (0.5)	0.33
Military Reserves and National Guard	-	-	-	-	-	-

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significant at α=0.1 level.
P-values with an asterisk (*) indicate a significant difference between the two rates.
An entry of '-' in a cell indicates that no or too few observations were available to meet statistical standards for reliability.

Table D-10. Class of Worker and Industry Consistency (in Percent): Mail

Category	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Private Households	20	93.7 (3.3)	38	98.0 (1.5)	-4.3 (3.6)	0.23
Not-for-profit organizations	57	84.3 (5.9)	66	80.5 (5.6)	3.9 (8.2)	0.64
Public Administration	196	92.8 (2.7)	211	94.7 (2.0)	-1.9 (3.6)	0.59
Postal Service	22	96.1 (3.0)	27	83.9 (9.3)	12.2 (9.9)	0.22
Active Duty Military	14	100.0 (0.0)	15	95.5 (3.8)	4.5 (3.8)	0.24
Military Reserves and National Guard	-	-	-	-	-	-

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significant at α=0.1 level.
P-values with an asterisk (*) indicate a significant difference between the two rates.
An entry of '-' in a cell indicates that no or too few observations were available to meet statistical standards for reliability.

Table D-11. Class of Worker and Industry Consistency (in Percent): CATI

Category	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Private Households	-	-	-	-	-	-
Not-for-profit organizations	-	-	-	-	-	-
Public Administration	36	100.0 (0.0)	58	98.9 (0.8)	1.1 (0.8)	0.17
Postal Service	-	-	-	-	-	-
Active Duty Military	-	-	-	-	-	-
Military Reserves and National Guard	-	-	-	-	-	-

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significant at α=0.1 level.
P-values with an asterisk (*) indicate a significant difference between the two rates.
An entry of '-' in a cell indicates that no or too few observations were available to meet statistical standards for reliability.

Table D-12. Class of Worker and Industry Consistency (in Percent): CAPI

Category	Test Sample Size	Test Percent	Control Sample Size	Control Percent	Test minus Control	P-Value
Private Households	60	100.0 (0.0)	52	99.0 (1.0)	1.0 (1.0)	0.33
Not-for-profit organizations	47	85.8 (6.5)	41	66.4 (10.1)	19.4 (12.8)	0.13
Public Administration	152	96.9 (1.3)	203	95.6 (1.6)	1.3 (2.1)	0.53
Postal Service	22	65.1 (17.2)	19	56.8 (15.9)	8.3 (23.9)	0.73
Active Duty Military	33	93.0 (6.8)	25	100.0 (0.0)	-7.0 (6.8)	0.31
Military Reserves and National Guard	-	-	-	-	-	-

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Standard errors are shown in parentheses. Significant at α=0.1 level.
P-values with an asterisk (*) indicate a significant difference between the two rates.
An entry of '-' in a cell indicates that no or too few observations were available to meet statistical standards for reliability.