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

September 19, 2017

2017 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS17-RER-11

MEMORANDUM FOR	Victoria Velkoff Chief, American Community Survey Office
From:	Karen Battle Chief, Population Division (POP)
Prepared by:	Jacquelyn Harth Population Division (POP)
Subject:	2016 American Community Survey Content Test Evaluation Report: Race and Hispanic Origin

Attached is the final report for the 2016 American Community Survey (ACS) Content Test for Race and Hispanic Origin. This report describes the results of the test for revised versions of the Race and Hispanic Origin questions.

If you have any questions about this report, please contact Jacquelyn Harth at 301-763-1530 or Angela Buchanan at 301-763-6040.

Attachment

cc:
Jennifer Ortman (ACSO)
Megan Rabe (ACSO)
Jennifer Reichert (ACSO)
Michael Bentley (DSSD)
Patrick Cantwell (DSSD)
Asaph Young Chun (DSSD)
Sarah Heimel (DSSD)
Elizabeth Poehler (DSSD)
Anthony Tersine (DSSD)
Nicholas Jones (POP)
Colleen Keating (POP)
Roberto Ramirez (POP)
Merarys Rios (POP)
Hyon Shin (POP)
Nicole Scanniello (SEHSD)

Intentionally Blank

September 19, 2017

2016 American Community Survey Content Test Evaluation Report: Race and Hispanic Origin

FINAL REPORT

Jacquelyn Harth, Angela Buchanan, Derek Breese, Merarys Rios, and Hyon B. Shin, Population Division
Sarah K. Heimel and Lindsay Longsine, Decennial Statistical Studies Division

Intentionally Blank

TABLE OF CONTENTS

EXECUTIVE SUMMARY vii
1 BACKGROUND 1
1.1 Justification for Inclusion of Race and Hispanic Origin in the Content Test 1
1.2 Question Development 3
1.3 Question Content 5
1.4 Research Questions 8
2 METHODOLOGY 9
2.1 Sample Design 9
2.2 Data Collection 10
2.3 Content Follow-Up 11
2.4 Analysis Metrics 11
2.4.1 Unit Response Rates and Demographic Profile of Responding Households 12
2.4.2 Item Missing Data Rates 13
2.4.3 Response Proportions 14
2.4.4 Benchmarks 15
2.4.5 Response Error 15
2.4.6 Other Analysis Methodology Specific to Race and Hispanic Origin 18
2.4.7 Standard Error Calculations 18
3 KEY RESEARCH FOR RACE AND HISPANIC ORIGIN 18
4 LIMITATIONS 18
5 RESEARCH QUESTIONS AND RESULTS 20
5.1 Unit Response Rates and Demographic Profile of Responding Households 20
5.1.1 Unit Response Rates for the Original 2016 ACS Content Test Interview 21
5.1.2 Unit Response Rates for the Content Follow-Up Interview 22
5.1.3 Demographic and Socioeconomic Profile of Responding Households 23
5.2 Item Missing Data Rates 25
5.3 Response Proportions 27
5.3.1 Race Category Response Proportions 27
5.3.2 Hispanic and MENA Reporting 30
5.4 Benchmarks 31
5.5 Response Error 32

5.6 Results for Analysis Specific to Race and Hispanic Origin 38
5.6.1 Detailed Reporting 38
5.6.2 Multiple-Response and Multiple-Race Reporting 44
5.6.3 Ancestry Information 46
5.6.4 Demographic Information 48
6 CONCLUSIONS 52
7 ACKNOWLEDGEMENTS 55
8 REFERENCES 55
Appendix A: Internet Versions of the Control and Test Questions 58
Appendix B: CATI/CFU and CAPI Versions of the Control and Test Questions 62
Appendix C: Unit Response Rates Supplemental Table 66

List of Tables

Table 1. Interview and Reinterview Counts For Each Response Category Used For Calculating the Gross Difference Rate and Index of Inconsistency 16
Table 2. Original Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode 21
Table 3. Mail Response Rates by Designated High (HRA) and Low (LRA) Response Areas 22
Table 4. Content Follow-up Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode of Original Interview 23
Table 5. Weighted and Unedited Response Distributions: Control Versus Test Treatment 24
Table 6. Comparison of Average Household Size 24
Table 7. Comparison of Language of Response 25
Table 8. Item Missing Data Rates for the Control Hispanic Origin Question and the Test Combined Question, by Mode 26
Table 9. Item Missing Data Rates for the Control Race Question and the Test Combined Question, by Mode 26
Table 10. Item Missing Data Rates for both Race and Hispanic Origin Questions, by Mode 27
Table 11. Overall Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race 27
Table 12. Mail Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race 28
Table 13. Internet Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race 29
Table 14. CATI Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race 29
Table 15. CAPI Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race 30
Table 16. Percent Reporting Hispanic Origin in Test and Control Treatments, by Mode 30
Table 17. Percent Reporting Middle Eastern or North African in Test and Control Treatments, by Mode 31
Table 18. Race and Ethnicity Percentages for the 2010 Census, the 2015 National Content Test (2015 NCT), and the American Community Survey Content Test (2016 ACS CT) 32
Table 19. Overall Difference in Gross Difference Rates (GDR) between Control and Test Treatments 33
Table 20. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – Mail 34
Table 21. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – Internet 34
Table 22. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – CATI 35
Table 23. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – CAPI 35
Table 24. Overall Difference in Index of Inconsistency (IOI) between Control and Test Treatments 36

Table 25. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – Mail 36
Table 26. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – Internet 37
Table 27. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – CATI 37
Table 28. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – CAPI 38
Table 29. Percent of All Persons Reporting a Detailed Race or Ethnicity Response, by Mode 39
Table 30. Percent of White Respondents Reporting a Detailed White Response (MENA Excluded), by Mode 40
Table 31. Percent of Hispanic Respondents Reporting a Detailed Hispanic Response, by Mode 40
Table 32. Percent of Black or African American Respondents Reporting a Detailed Black or African American Response, by Mode 41
Table 33. Percent of Asian Respondents Reporting a Detailed Asian Response, by Mode 41
Table 34. Percent of American Indian or Alaska Native (AIAN) Respondents Reporting a Detailed AIAN Response, by Mode 42
Table 35. Percent of Middle Eastern or North African (MENA) Respondents Reporting a Detailed MENA Response, by Mode 42
Table 36. Percent of Native Hawaiian or Other Pacific Islander (NHPI) Respondents Reporting a Detailed NHPI Response, by Mode 43
Table 37. Percent of Some Other Race or Ethnicity (SOR) Respondents Reporting a Detailed SOR Response, by Mode 43
Table 38. Multiple-Response Reporting for Control and Test Treatments, by Mode 44
Table 39. Percent of Hispanic Respondents Reporting a Race 45
Table 40. Multiple-Race Reporting for Control and Test Treatments, by Mode 45
Table 41. Percent of Respondents with Consistent Race or Ethnicity and Ancestry Responses for Control and Test Treatments, by Mode 47
Table 42. Ancestry Item Missing Data Rates for Control and Test Treatments, by Mode 47
Table 43. Race and Ethnicity Reporting in Test and Control Treatments, by Age 49
Table 44. Race and Ethnicity Reporting in Test and Control Treatments, by Sex 50
Table 45. Race and Ethnicity Reporting in Test and Control Treatments, by Educational Attainment (Persons 25 and Older) 51
Table 46. Race and Ethnicity Reporting in Test and Control Treatments, by Housing Tenure 52
Table C-1. Unit Response Rates by Designated High (HRA) and Low (LRA) Response Areas. 66

List of Figures

Figure 1. Control Version of Race and Hispanic Origin Questions (Mail) 6
Figure 2. Test Version of Race and Hispanic Origin Question (Mail) 7
Figure 3. Ancestry Question 46
Figure A-1. Internet Version of the Control Treatment Hispanic Origin Question 58
Figure A-2. Internet Version of the Control Treatment Race Question 58
Figure A-3. Image of the Internet Version of the Test Treatment Combined Race and Hispanic Origin Question 59
Figure A-4. White Detailed Question for the Test Treatment Race and Hispanic Origin Question 59
Figure A-5. Hispanic Detailed Question for the Test Treatment Race and Hispanic Origin Question 59
Figure A-6. Black Detailed Question for the Test Treatment Race and Hispanic Origin Question 60
Figure A-7. Asian Detailed Question for the Test Treatment Race and Hispanic Origin Question 60
Figure A-8. AIAN Detailed Question for the Test Treatment Race and Hispanic Origin Question 60
Figure A-9. MENA Detailed Question for the Test Treatment Race and Hispanic Origin Question 61
Figure A-10. NHPI Detailed Question for the Test Treatment Race and Hispanic Origin Question 61
Figure A-11. SOR Detailed Question for the Test Treatment Race and Hispanic Origin Question 61

Intentionally Blank

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 with 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 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 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 (CATI) period for nonresponse follow-up; and 3) a one-month Computer-Assisted Personal Interview (CAPI) period for a sample of the remaining nonresponse. For housing units that completed the original 2016 ACS Content Test interview, a Content Follow-Up telephone reinterview was conducted to measure response error.

Race and Hispanic Origin

Over the past decade, the Census Bureau has implemented research to address known issues with race and Hispanic origin reporting and concerns raised by data users and community organizations. Efforts to improve the quality of data collected from the race and Hispanic origin questions focused on:

- Improving accuracy and reliability of reporting in the major race and ethnicity categories as defined by the U.S. Office of Management and Budget (OMB),[1]
- Collecting detailed data for all major groups, and
- Obtaining lower item nonresponse rates.

Prior to the 2010 Census, Census Bureau researchers identified potential issues with the collection of data on race and Hispanic origin, such as a growing number of respondents who chose the response option of Some Other Race (SOR), which was meant to be a residual category. The Census Bureau began to investigate these issues starting with the 2010 Census

[1] The major race groups as defined by the Office of Management and Budget's *1997 Revisions to the Standards for the Classification of Federal Data* (U.S. Office of Management and Budget, 1997) are White, Black or African American, Asian, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander. The major ethnicity categories are Hispanic or Latino, and not Hispanic or Latino.

Race and Hispanic Origin Alternative Questionnaire Experiment (2010 AQE). The 2010 AQE showed the increase of SOR responses was due primarily to Hispanic respondents who, when answering the race question, did not identify with the listed races and instead chose SOR. Another finding from the 2010 AQE was that many people of Middle Eastern or North African (MENA) origin did not identify with any of the existing categories.

In 2015, the Census Bureau conducted the 2015 National Content Test (NCT) to determine which question format and wording would elicit the best quality data on race and ethnicity.[2] In addition to testing different question format approaches, alternative instructions, and terminology, the 2015 NCT tested a distinct category for respondents of MENA origin, which is not a minimum reporting category within the current OMB standards.[3]

While the optimal design for collecting data on race and ethnicity was determined by the 2015 NCT, the 2016 ACS Content Test served as an operational test of the concepts that were investigated in the 2015 NCT. The 2016 ACS Content Test provided an opportunity to test additional data collection modes and to examine contextual data from the ACS characteristic variables. Specifically, the 2016 ACS Content Test evaluated interviewer-administered collection modes, assessed the race and ethnicity questions against demographic and socioeconomic data, and separately compared the race and ethnicity results to data from the ancestry question.

Half of the sample was assigned to the control treatment, which asked separate Hispanic origin and race questions without a MENA category, and the other half was assigned to the test treatment, which asked a combined Hispanic origin and race question with a distinct MENA category. In both the control and test treatments, detailed races and ethnicities were collected from all categories through either a write-in line or a checkbox. Within each treatment, the internet, mail, CATI, and CAPI modes were all designed similarly, except that the internet mode design on the test treatment provided six checkboxes to collect detailed race and ethnicities from all major groups.

The results of this test will not be used to make a recommendation about the question format for race and Hispanic origin; rather, they will be used to inform decision makers about whether it is feasible to collect a combined race and ethnicity question in all modes of the ACS.

[2] In this report, the term "race and ethnicity" is often used interchangeably with "race and Hispanic origin."

[3] The Census Bureau has been working with the OMB and other federal statistical agencies over the past several years to explore ways for improving data on race and ethnicity. In September 2016, the OMB issued a Federal Register Notice (FRN) to advise the public that it would be undertaking a limited review of the federal standards on the collection of race and ethnicity data, and that they may issue subsequent notices to elicit public comment. A second OMB FRN was issued on March 1, 2017 for public comment on the race and ethnicity standards. It is expected that the OMB and federal agencies will continue to engage the public and stakeholders as these discussions continue. The OMB has advised that it will make a decision on the standards by the end of 2017.

Results

The analyses of the race and Hispanic origin questions were guided by several research questions. A summary of the results is provided below:

- The proportions for each major race and ethnicity group generally reflected the distributions previously observed in the 2010 Census as well as the 2015 NCT.[4]
- Item missing data rates, which represent a respondent's understanding and willingness to answer questions, were lower for the test treatment when compared with the race question in the control treatment. When comparing the item missing data rates in the test treatment to both the race and Hispanic origin questions in the control treatment, no significant differences were detected.
- The reporting of White and SOR responses was lower in the test treatment compared with the control treatment, and this result was consistent across data collection modes, with the exception of White responses in the CATI mode. American Indiana and Alaska Native (AIAN) responses were also significantly lower in the test treatment compared with the control treatment overall and in the CAPI mode. There were no significant differences between treatments for the Black, Asian, and Native Hawaiian and Other Pacific Islander (NHPI) categories.[5]
- There were no significant differences in the proportion of Hispanic responses between the treatments. Likewise, there were no significant differences in the proportion of MENA responses between the treatments.[6]
- Overall, the reporting of detailed responses, which occurred when a respondent provided a more detailed race or ethnic identity beyond the major OMB group (i.e., reporting Irish as part of a White identity, Cuban as part of a Hispanic identity, or Chinese as part of an Asian identity), was higher in the test treatment than in the control treatment. This held true for respondents identifying as White, Black, or SOR. The control treatment did not offer detailed reporting checkboxes for White and Black groups, whereas the test treatment had detailed checkboxes in the internet mode. There were no significant differences in detailed reporting for AIAN, MENA, or NHPI. For Hispanic and Asian, the proportion reporting detailed responses was higher in the control than in the test. The control treatment offered detailed reporting checkboxes for Hispanic and Asian groups in all modes, whereas the test treatment only had detailed checkboxes in the internet mode.
- Multiple-response reporting, or a respondent reporting Hispanic origin in addition to a race, was lower in the test, due to a large proportion of Hispanic respondents reporting Hispanic as their only identity. Multiple-race reporting, or when a respondent reported two or more races, was also lower in the test treatment.
- Response reliability results, which compared reinterview responses on race and ethnicity with original responses given by the same person, indicated that the test treatment was more reliable for those who reported as White. There were no significant differences in reliability for those who reported as Hispanic, Black, Asian, AIAN and NHPI.

---

[4] Race reporting represents the population who indicated a race alone or in combination.

[5] Alaska and Hawaii were not included in the sample for the 2016 ACS Content Test. Therefore, the results for AIAN and NHPI should be interpreted with caution.

[6] Hispanic and MENA reporting represent the population who indicated Hispanic alone or in combination, or MENA alone or in combination.

- Overall, respondents in the test treatment reported their race and Hispanic origin more consistently with ancestry than those in the control treatment. Responses were considered consistent for a respondent when their race and Hispanic origin responses exactly matched their ancestry response. In examining the results by mode, the mail test treatment had significantly higher consistency in race or ethnicity reporting compared with the control. In contrast, the CATI mode had significantly lower response consistency in race or ethnicity reporting in the test treatment.
- For ancestry, the overall item missing data rate was higher for the test treatment compared with the control. Item missing rates for ancestry in the internet and CAPI modes were also higher in the test treatment than in the control treatment.
- For all demographic and socioeconomic groups that we examined, White and SOR responses were lower in the test treatment compared with the control. This result was consistent across modes.

Conclusions

Prior research has shown that Hispanic respondents, when answering a combined question, tend to report as being only Hispanic and are less likely to report being White or SOR. Thus, the finding that White and SOR responses were lower in the test treatment was not surprising. This is more in line with how Hispanics self-identify with respect to race and ethnicity; the level of White reporting for the test treatment was similar to the level of non-Hispanic White reporting observed with the traditional two separate questions approach. The importance of this finding is two-fold. First, in a combined question format, many Hispanic respondents identified as being Hispanic without any indication of being any other race or ethnicity. Second, the lower response in SOR was due mostly to Hispanics no longer choosing that category when asked to choose a race. The 2010 AQE and the 2015 NCT also demonstrated that with a combined question format, respondents used the SOR category as the residual response option that it was originally intended to be. Finally, the other race and ethnicity categories appeared to be minimally impacted by the test treatment. There were no significant differences between treatments for the Hispanic, Black, Asian, and NHPI categories.

Taken together, the 2016 ACS Content Test results for race and ethnicity confirmed the results from the 2010 AQE and the 2015 NCT in that a combined question format and use of the MENA category results in higher data quality for race and ethnicity. Additionally, the 2016 ACS Content Test indicated that quality race and ethnicity data can be collected in the ACS environment using a combined question format and MENA category.

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 testing and results for Race and Hispanic Origin.

1.1 Justification for Inclusion of Race and Hispanic Origin in the Content Test

A question about race has been asked on the census since 1790 and data on Hispanic origin have been collected since 1970. The categories used and the information collected has changed over time, reflecting changes in society and shifts in the way the Census Bureau classifies race and ethnicity (Pratt et al., 2015). The current race and ethnicity questions are based on the U.S. Office of Management and Budget's (OMB) *1997 Revisions to the Standards for the Classification of Federal Data* (U.S. Office of Management and Budget, 1997). All federal statistical agencies, including the Census Bureau, are required to adhere to the OMB standards when collecting data on race and ethnicity.

While many respondents report within the race and ethnicity categories specified by the OMB 1997 standards, it is clear from recent censuses, surveys, and experimental tests that the response options are not well understood or accepted by a growing number of respondents (Compton et al., 2013). As a result, some respondents are unable or unwilling to self-identify in the categories provided by the OMB standards. For a segment of respondents, this arises because of the

conceptual complexity rooted in the OMB standards' distinction between "race" and "ethnicity," as well as the format of the current race and Hispanic origin questions and categories.

The growing lack of understanding or acceptance of the 1997 OMB standards is compounded by the rapidly changing demographics of the U.S. population. The increasing racial/ethnic diversity and complexity of our nation's population reflects increases in interracial marriage and multiracial children, immigration flows from all corners of the globe, and racial and ethnic self-identification reporting patterns that are fluid over time. As a result, there are an increasing number of responses of Some Other Race (SOR) and widespread campaigns by advocacy groups wanting changes to the race and Hispanic origin questions that reflect the need for representative categories. All of these realities heighten the need to explore alternatives to the current format for collecting data on race and ethnicity.

The 2010 Census Alternative Questionnaire Experiment (AQE) tested several versions of a combined race and Hispanic origin question and changes to the two separate questions. The combined question approach demonstrated lower item nonresponse and a significant reduction in the proportion of the population classified as "Some other race" compared with the two separate questions approach. Additionally, the 2010 AQE demonstrated that a combined question approach has the impact of both Hispanics and non-Hispanics alike finding a place to successfully identify and report their race and/or origin. For example, based on the results of the 2010 AQE content reinterview, the combined question better captured self-identity for Whites than the two separate questions approach (Compton et al., 2013).

Focus group research from the 2010 AQE found that many people with a Middle Eastern or North African (MENA) background do not consider themselves White and were unsure about what category or categories to select. Under current OMB standards, the White category is defined as including people with origins from Europe, the Middle East, and North Africa. However, when the revised 1997 OMB standards were published, they called for more research into how people with Arab or Middle Eastern heritage should be classified. In 2013, the Census Bureau received a formal request from the Arab American Institute and 25 other organizations and community members to add a distinct category to the 2020 Census for people with MENA origins. Feedback from organizational leaders and scholars who attended a Census Bureau forum on the topic in 2015, as well as thousands of comments to a Federal Register Notice announcing the Census Bureau's plans to test a MENA category as part of the 2015 National Content Test (NCT), supported the testing of a MENA category (Buchanan et al., 2016).

The 2015 NCT had a robust sample of 1.2 million housing unit addresses and included oversampling for key race and ethnicity groups. This, coupled with the implementation of a complex test design with eight different paper question designs as well as a fully factorial design with 36 different internet versions, enabled the NCT to undertake the complex and important challenge of determining the optimal designs that can be used for collecting Census race/ethnicity data. The 2015 NCT examined several key dimensions for improving the data on race and ethnicity, including question format (separate versus combined approach), response categories (inclusion of a distinct MENA category), instruction wording, and question terminology. The 2015 NCT also included a reinterview component in which households who

responded to the initial survey were reinterviewed with a detailed set of questions to ascertain the "truth" of their self-reported race and ethnicity answers.[7]

The 2016 ACS Content Test presented an important opportunity to further test a combined race and Hispanic origin question, as a complement to the 2015 NCT, since question changes that are being considered for the 2020 Census are also being considered for the ACS. The 2016 ACS Content Test also allowed for the testing of an enhanced version of the traditional two separate questions approach, using multiple modes in both English and Spanish. The research for the 2016 ACS Content Test built upon the results from the 2010 AQE, which included testing combined questions and modified separate questions (Compton et al., 2013); preparation for the 2016 ACS Content Test involved close collaboration with the preparation for the 2015 NCT. Specifically, cognitive testing conducted in preparation for the 2016 ACS Content Test helped to inform the design of the question wording included in both tests.

While the design of the 2015 NCT allowed for more robust treatments of question revisions than the two treatments in the 2016 ACS Content Test, the 2016 ACS Content Test provided additional questionnaire design and data mode research opportunities, including an evaluation of the performance of the distinct MENA category in interviewer-assisted modes. For example, while the 2015 NCT was designed for self-response modes (internet, paper, telephone questionnaire assistance), the 2016 ACS Content Test allowed for additional testing in Computer-Assisted Telephone Interview (CATI) and Computer-Assisted Personal Interview (CAPI) modes. The 2016 ACS Content Test also provided an opportunity to examine selected socioeconomic characteristics in connection with responses to the modified questions. Furthermore, the 2016 ACS Content Test provided the opportunity to evaluate the impact of the combined race and Hispanic origin question on the ancestry question used in production ACS. The ancestry question is an open-ended question asked later in the survey that asks what a person's "ancestry or ethnic origin" is (see Section 5.6.3 for more information on ancestry).

1.2 Question Development

Initial versions of the new and revised questions for the 2016 ACS Content Test were proposed by federal agencies participating in the 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.[8] 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

[7] Note that self-identified race and Hispanic origin can change over circumstances and time; it does not have an absolute "truth." However, the conceptual goal of the 2015 NCT reinterview was to get closer to how a respondent would typically self-identify his/her race or origin, through a series of more detailed questions and probes.

[8] 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 the OMB, as well as the Interagency Council on Statistical Policy Subcommittee, reviewed these proposals for the ACS. The 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 ACS Content Test (Stapleton and 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 in English and Spanish for one version of the questionnaire 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., 2015). 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.[9] 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.[10]

Specific to the development of the questions related to race and Hispanic origin, the first round of cognitive testing employed both a separate (two questions) version and a combined (one question) version. The findings showed a strong preference for a combined version, a positive reaction to the addition of a category for Middle Eastern or North African, and wide variation in respondents' interpretations of the terms "race" and "origin."

The second round of cognitive testing sought to further test the performance of a combined question, using modified wording that eliminated any reference to race, ethnicity, or origin in the question text, instead asking respondents which "categories" described them. As in the first round, respondents were asked to provide details about the category they selected to understand more about their backgrounds. Cognitive probes were designed to ascertain whether respondents understood what the overall question was asking, if the response categories reflected how respondents thought of themselves, and whether asking for details about the categories chosen led them to think about race, ethnicity, ancestry, place of birth, or some other classification.

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

[10] 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.

In the third round of cognitive testing, most group quarters respondents understood the intent of the combined race and Hispanic origin question and were able to provide an answer that fit how they think of themselves. The problems observed seemed to be arbitrary. The third round of cognitive testing also included respondents from Puerto Rico, and the issues observed in the first two rounds appeared to be resolved by wording changes following the second round. Therefore, no further revisions to the combined race and Hispanic origin question were recommended based on the findings from round three.

Cognitive and usability testing was also conducted in conjunction with the 2015 NCT (Meyers et al., 2015; Sha et al., 2016). Results from both sets of qualitative research were used to inform design decisions for both the 2015 NCT and the 2016 ACS Content Test.

1.3 Question Content

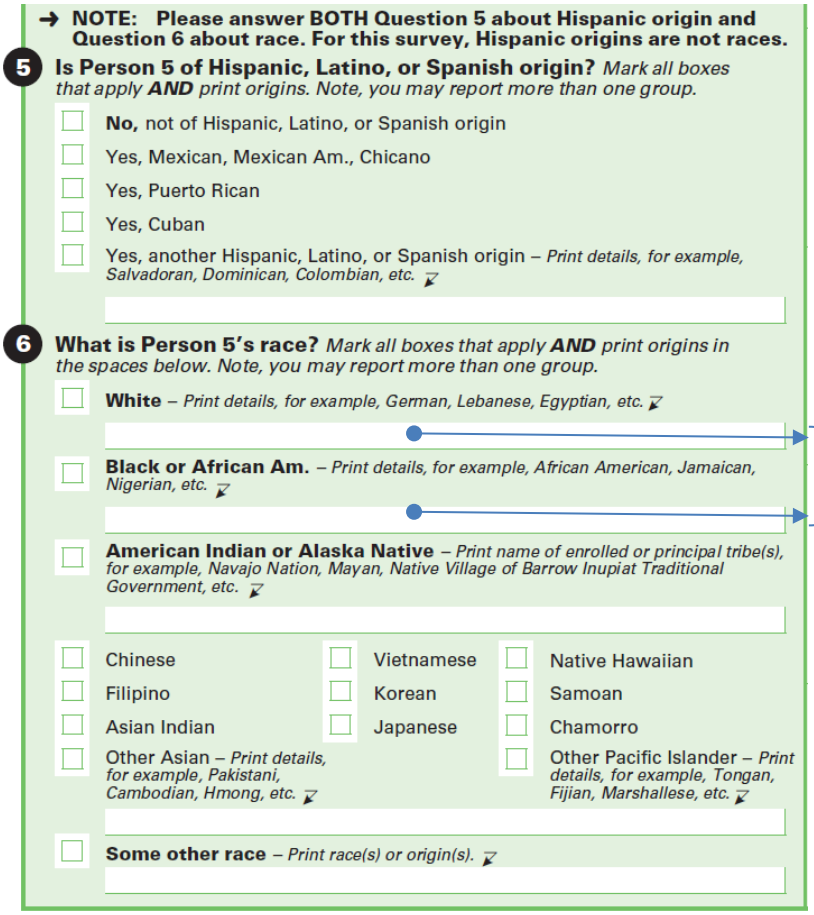
The control treatment format of the question was not the current version used in production ACS, but an alternative two-question version with write-in areas added for the White and Black categories. The test treatment format asked which "categories" described the respondent, and combined the Hispanic origin and race categories into one question, with "Hispanic, Latino or Spanish origin" as one of the response options. The total number of detailed race and ethnicity checkbox response options was reduced in the mail response mode for the test treatment, but a write-in response area designed to elicit more detailed information about each person's racial and ethnic origins was available for all response categories. The internet response mode allowed for more flexibility in the question design, expanding the total number of detailed check-box response options in the test treatment for all race and ethnicity categories.

A distinct response option was also added for MENA responses in all modes in the test treatment. This new category was intended to collect detailed responses for both nationalities and transnational ethnicities from the Middle East and North Africa region, such as Lebanese, Egyptian, and Kurdish.[11]

Figures 1 and 2 show the control and test versions of each question, respectively, as they appeared on the paper questionnaire that was mailed to sampled households. Interviewer-assisted versions of the questionnaire had the same content formatted according to mode (see Appendix B for the CATI and the CAPI versions). Note that the paper questionnaire listed three examples for each major OMB group due to space constraints, while the CATI and CAPI modes listed six examples, as was designed in the 2015 NCT. The internet version of the questions had additional checkboxes for the six detailed groups, as well as three more examples above the write-in line (see Appendix A for the internet versions). Groups were chosen to be examples based on their population size in the 2010 Census and the 2013 ACS.

[11] The working classification of the MENA category includes: Algerian, Bahraini, Egyptian, Emirati, Iraqi, Iranian, Israeli, Jordanian, Kuwaiti, Lebanese, Libyan, Moroccan, Omani, Palestinian, Qatari, Saudi Arabian, Syrian, Tunisian, Yemeni, Amazigh or Berber, Assyrian, Bedouin, Chaldean, Copt, Druze, Kurdish, and Syriac. For more information on how the MENA category and classification were developed, see the MENA Forum page at: https://www.census.gov/library/working-papers/2015/demo/2015-MENA-Experts.html.

Figure 1. Control Version of Race and Hispanic Origin Questions (Mail)



➜ **NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this survey, Hispanic origins are not races.**

5 **Is Person 5 of Hispanic, Latino, or Spanish origin?** *Mark all boxes that apply **AND** print origins. Note, you may report more than one group.*

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin – *Print details, for example, Salvadoran, Dominican, Colombian, etc.*

6 **What is Person 5's race?** *Mark all boxes that apply **AND** print origins in the spaces below. Note, you may report more than one group.*

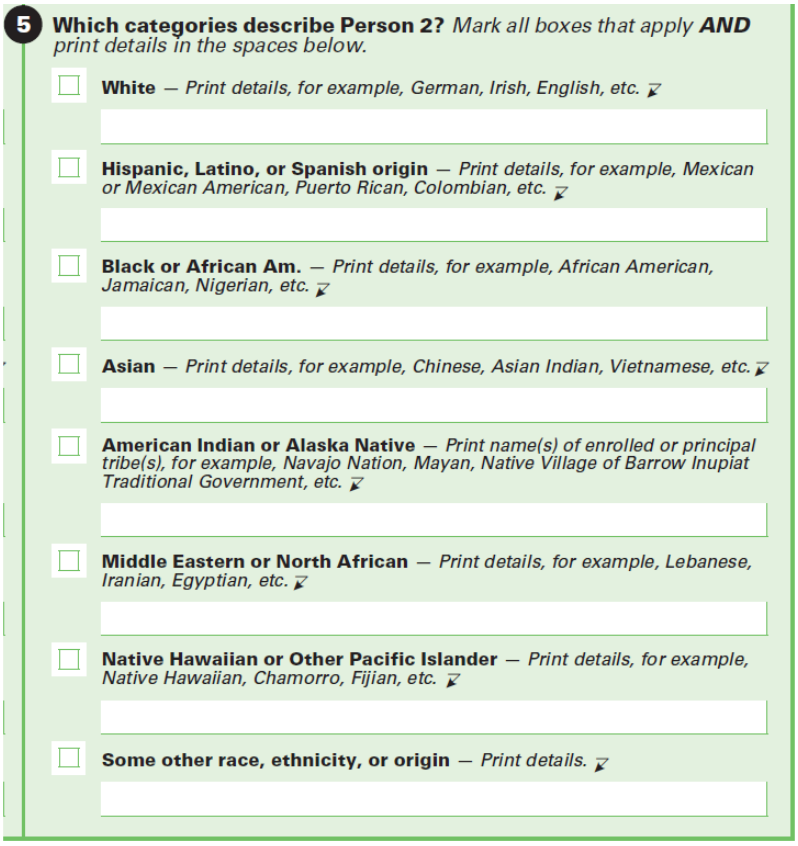
- ☐ **White** – *Print details, for example, German, Lebanese, Egyptian, etc.*
- ☐ **Black or African Am.** – *Print details, for example, African American, Jamaican, Nigerian, etc.*
- ☐ **American Indian or Alaska Native** – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Mayan, Native Village of Barrow Inupiat Traditional Government, etc.*

- ☐ Chinese
- ☐ Filipino
- ☐ Asian Indian
- ☐ Vietnamese
- ☐ Korean
- ☐ Japanese
- ☐ Native Hawaiian
- ☐ Samoan
- ☐ Chamorro
- ☐ Other Asian – *Print details, for example, Pakistani, Cambodian, Hmong, etc.*
- ☐ Other Pacific Islander – *Print details, for example, Tongan, Fijian, Marshallese, etc.*

- ☐ **Some other race** – *Print race(s) or origin(s).*

These two write-in spaces were not present in the 2015 ACS production version of the questionnaire.

Figure 2. Test Version of Race and Hispanic Origin Question (Mail)



The key differences between the control and test questions were in question format, response categories, terminology, and internet mode content. In the control version, two separate questions were used for collecting data on race and Hispanic origin. In the test version, the question format differed in that the two separate questions were combined into one. Additionally, the control version did not include a MENA response category and the MENA examples were listed with the White category. The test version contained a distinct MENA category and relevant examples (MENA examples were removed from the White category). With respect to terminology, the differences between the two versions were that the control version used the terms "race" and "origin(s)" in the question stems, while the test version avoided using the terms "race," "ethnicity," and "origin" in the question stem and instead used the term "categories."

In the test treatment, the internet mode provided a series of additional screens to collect data on all major groups. The major groups were White, Hispanic, Black, Asian, American Indian or Alaska Native (AIAN), MENA, Native Hawaiian or Other Pacific Islander (NHPI), and SOR. On the initial screen, the instrument collected data on the major groups via a checkbox and examples (see Appendix A). The examples provided were the six largest detailed groups residing in the United States that represented the geographic diversity of the race or ethnicity group's definition. For any selected group from the initial screen, a subsequent screen was utilized to gather more details. For instance, if a respondent indicated they were White and Asian, the subsequent screen allowed the respondent to check up to six detailed White groups (such as German) and also allowed more detail to be provided on a write-in line. Next, a similar screen

allowed the respondent to see the detailed Asian checkboxes (such as Chinese) and allowed for additional write-ins.

1.4 Research Questions

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

The following research questions were examined:[12]

1. *How do the estimated proportions of persons in each race or ethnic category for each treatment (control and test) compare with the proportions from the 2010 Census, the 2010 AQE, and the 2015 NCT?*

2. *Which question version has lower item missing data rates overall and by mode?*

3. *How do the estimated proportions of persons in each of the six major race groups compare between control and test versions overall and by mode?*

4. *How does detailed reporting for all race or ethnic categories (including Hispanic and MENA) proportionally differ overall and by mode?*

5. *How do the estimates for the overall Hispanic origin and MENA populations compare between the control and test versions overall and by mode?*

6. *How do the proportions of multiple-response and multiple-race persons compare between control and test versions overall and by mode?*

7. *Using results from CFU, which question version results in estimates that are more reliable overall and by mode?*

8. *How does the information obtained from the race and Hispanic origin question write-ins in the control version compare with the information obtained from the ancestry question overall and by mode? How does information obtained from the race and Hispanic origin question write-ins in the test version compare with information obtained from the ancestry question overall and by mode?*

[12] An additional research question "How do the estimates of the overall White population compare between the control and test versions overall and by mode?" was included in the ACS Research & Evaluation Analysis Plan for Race and Hispanic Origin but was not included in this report, as it was already covered in the third research question.

9. *Does the item missing data rate for ancestry differ between control and test versions overall and by mode?*

10. *How do the estimates of race and Hispanic origin crossed by socioeconomic and demographic characteristics differ 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 2016 ACS Content Test sample universe did not include GQs, nor did it include housing units in Alaska, Hawaii, or Puerto Rico.[13] The sample design for the 2016 ACS Content Test was largely based on the ACS production sample design with some modifications to better meet the test objectives.[14] 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.[15] 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.[16] For the 2016 ACS 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 2016 ACS Content Test was adjusted to meet the budget constraint.

---

[13] 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.

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

[15] 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.

[16] 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 2016 ACS 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 2016 ACS Content Test cases based on their geographic proximity to the cases and therefore could also conduct both control and test cases.

The 2016 ACS Content Test's data collection protocol differed from the production ACS in a few significant ways. The 2016 ACS 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 2016 ACS Content Test excluded the telephone Failed Edit Follow-Up (FEFU) operation.[17] Furthermore, the 2016 ACS Content Test had an additional telephone reinterview operation used to measure response reliability. This telephone reinterview component is referred to 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 2016 ACS Content Test had Spanish-language automated instruments; however, there were no paper versions of the 2016 ACS Content Test questionnaires in Spanish.[18] Any case in the 2016 ACS Content Test sample that completed a Spanish-language internet, CATI, or CAPI response was included in the analysis. However, if a case sampled for the 2016 ACS 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 2016 ACS 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 2016 ACS Content Test. CFU interviews for the 2016 ACS Content Test were conducted in either Spanish or English. The practical need to limit the language response

[17] 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.

[18] 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.

options for 2016 ACS Content Test respondents is a limitation to the research, as some respondents self-selected out of the 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.[19] 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.[20] 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.[21]

2.4 Analysis Metrics

This section describes the metrics used to assess the revised version of the race and Hispanic origin question. Those metrics include the item missing data rate, response distributions, comparisons to benchmarks, response error, and other metrics. This section also describes the methodology used to calculate unit response rates and standard errors for the test.

All 2016 ACS 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

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

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

[21] Refer to the 2016 ACS Content Test report on Retirement Income for a discussion on CFU questions for survivor, disability, and retirement income.

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 2016 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 $\alpha = 0.1$; all hypothesis tests in this report are two-tailed. Since numerous comparisons were conducted 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).

2.4.1 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.[22] 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 2016 ACS 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 2016 ACS Content Test topics was done by high and low response areas. Using the 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 consisted 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 are 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

[22] 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.

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 2.4.5.) relies upon CFU interview data. Statistical differences between CFU response rates for control and test treatments was not taken as evidence that one version was 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 2016 ACS 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 2016 ACS 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.[23]

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 significantly 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).

2.4.2 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

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

is the proportion of eligible units (housing units for household-level items or persons for person-level items) for which a required response is missing.

Analysis of the control treatment used three item missing data rates: one strictly for the Hispanic origin question, one strictly for the race question, and one for the combination of both the Hispanic origin and race questions (when a response is missing for both the Hispanic origin and race questions). The test treatment had one item missing data rate for the entire race and Hispanic origin question.

An initial comparison was made between the item missing data rate from the test treatment to the item missing data rate from only the Hispanic origin question from the control treatment. A second comparison was made between the item missing data rate from the test treatment to the rate from only the race question from the control treatment. A third comparison was made between the item missing data rate from the test treatment to the rate of those who have missing values for both the Hispanic origin and race questions on the control treatment. Person records were considered to have missing data if they either had no response to any part of the question (completely blank), answered "Don't know" or "Refused", or provided only an uncodable write-in response.

The following guidelines were made for this analysis:

- Absence of a write-in given the presence of a checkmark did not constitute missing data for the purpose of this question.
- For the test treatment, an uncodable response was defined as one where only an uncodable write-in response was provided. For example, a respondent did not mark any checkboxes and wrote in "I am from Mars."
- For the control treatment, an uncodable response was defined as one where either only uncodable write-in responses were provided in both the Hispanic origin and race questions, or only uncodable write-in responses were provided in one question and nothing was provided in the other question. For example, a respondent wrote "I am from Mars" on both the Hispanic origin and race questions or a respondent left the Hispanic origin question blank and then wrote "I am from Mars" in response to the race question (or vice versa).
- Persons who provided an uncodable write-in but also marked a checkbox were not counted as 'uncodable' for this analysis. For example, a respondent marked the Black checkbox, but wrote, "I am from Mars." This precluded any cases from automated instruments in the test treatment being identified as uncodable, as well as cases from the control CATI and CAPI interviews.

2.4.3 Response Proportions

Comparing the response proportions between the control version of a question and the test version of a question allowed us to assess whether the question change affected the resulting estimates.

Proportion estimates were calculated as:

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

2.4.4 Benchmarks

For the topic of race and Hispanic origin, data from both the control and test treatments were compared with information from the 2010 Census and the 2015 NCT. Due to the methodological differences among the 2010 Census, the 2015 NCT, and the 2016 ACS Content Test, statistical testing was not conducted. Instead, the data from the 2010 Census and the 2015 NCT are presented as a reference.

Note that the questions in the 2016 ACS Content Test differed from the 2010 Census questions in that there were examples and write-in areas for the White and Black categories on the race question that were not in the 2010 Census. Additionally, 2010 Census data were edited or imputed for respondents who did not provide a race or Hispanic origin response, while the 2016 ACS Content Test data were not. Also note that the test treatment data in the 2016 ACS Content Test for the MENA population could not be compared with the 2010 Census data since MENA was not a response option in the 2010 Census. According to the OMB standards, people who reported MENA origins in the 2010 Census were classified as White.

The 2015 NCT was a multi-panel test that consisted of different treatment paths for question format (separate or combined), inclusion of a MENA category, wording of instructions, and question terminology. Therefore, the data presented in Table 18 for the 2015 NCT separate and combined questions are exclusively from Panel 1 (separate format) and Panel 36 (combined format), as these panels were the closest approximation to the content of the 2016 ACS Content Test control and test treatments, respectively.

The 2010 AQE analysis only presented results for the population reporting each race and ethnicity category alone, whereas the 2010 Census and the 2015 NCT presented results for the population alone or in combination with another category. The 2010 AQE also did not include a MENA category. Therefore, despite its inclusion in the research questions, we did not present the race and ethnicity distributions for the 2010 AQE in this report.

2.4.5 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.[24] 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, the 2010 AQE, and the 1990, 2000, and 2010 decennial censuses (Dusch & Meier, 2012).

The following measures were calculated for individual response categories and were used to evaluate consistency:

- Gross difference rate (GDR)
- Index of inconsistency (IOI)

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

Table 1. 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

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

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

[24] A majority of the CFU interviews were conducted with the same respondent as in the original interview (see the Limitations section for more information). For the race and ethnicity results, the analysis was limited to only cases that had the same respondent.

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:

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

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.

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 1 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:

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

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

2.4.6 Other Analysis Methodology Specific to Race and Hispanic Origin

The race and ethnicity data were coded differently for the 2016 ACS Content Test than in production ACS. Both the control and test versions of the race and ethnicity question(s) allowed respondents to write in White (including MENA) and Black ethnic groups. The 2015 NCT code list was used, which provided an expanded number of groups that could be coded within the White, Black, and MENA categories. We also allowed up to ten groups to be coded from each write-in line. In production ACS, only one Hispanic origin code is allowed and up to eight race groups are recorded, with a maximum of two from each write-in line. A new auto-coding file was used that allowed the wider range of groups to be coded for the 2016 ACS Content Test.

2.4.7 Standard Error Calculations

We estimated the variances of the estimates using the successive differences replication 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:

$$\mathrm{Var}(\mathrm{X}_0) = \frac{4}{80}\sum_{\mathrm{r}=1}^{80}(\mathrm{X}_\mathrm{r} - \mathrm{X}_0)^2$$

where:

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

3 KEY RESEARCH FOR RACE AND HISPANIC ORIGIN

Inclusion of race and Hispanic origin in the 2016 ACS Content Test allowed for evaluation of estimates of race and Hispanic origin crossed by selected socioeconomic and demographic characteristics, as well as operationalization of coding for these question versions. Additionally, certain research questions were answered by mode, as previously indicated, in order to observe any differences between the control and test treatments in CATI and CAPI modes and the self-response modes.

4 LIMITATIONS

CATI and CAPI interviewers were assigned control and test treatment cases, as well as production cases. The potential risk of this approach was the possible introduction of a cross-contamination or carry-over effect due to the same interviewer administering multiple versions of the same question item. Interviewers were trained to read the questions verbatim to minimize this risk, but there still exists the possibility that an interviewer may have deviated from the

scripted wording of one question version to another. This could potentially mask a treatment effect from the data collected.

The 2016 ACS Content Test 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 because that topic was tested as part of the 2016 ACS Content Test.

The CFU reinterview was not conducted in the same mode of data collection for households that responded by internet, mail, or 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, a different member of the household was interviewed 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.[25] The difference between the test and control treatments is not statistically significant (p-value = 0.26). The cases in which there was a different respondent were dropped from the race and ethnicity analysis of the reliability metrics. This may have introduced a potential bias if the excluded cases were somehow systematically different from the retained cases.

The 2016 ACS Content Test did 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 with 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 2016 ACS Content Test sample

[25] 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.

design. Therefore, workload and budget estimates were too low. In order to stay within the budget, the CAPI 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 2016 ACS Content Test cases to be mailed production ACS questionnaires instead of 2016 ACS Content Test questionnaires. There were 49 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 was 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 had the potential to impact the response reliability results.

The assumption of parallel measures for the GDR and IOI calculations was not met for the following categories: White (overall), Hispanic (overall), and SOR (overall, internet mode, and CAPI mode) (see Appendix C for Net Difference Rates). For these categories, the GDR and IOI estimates are biased to some extent.

There were some experimental design limitations to the 2016 ACS Content Test for race and Hispanic origin. First, the control did not mimic the production ACS questions on race and Hispanic origin. The control treatment allowed all respondents to provide detailed ethnicities, whereas on the production ACS this was not available for the White and Black categories. Additionally, the test version differed from the control, not only in format (combined question), but also in instruction wording. Therefore, it may be difficult to tell which change caused any differences in reporting. Also, the absence of Alaska and Hawaii in the sample may have resulted in underrepresentation of the AIAN and NHPI populations; therefore, the results for these groups should be interpreted with caution.

5 RESEARCH QUESTIONS AND RESULTS

This section presents the results from the analyses of the 2016 ACS Content Test data for the race and Hispanic origin question(s). 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 2016 ACS Content Test interview and 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 2016 ACS 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 2 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. Looking at the overall unit response rate (all modes combined), the difference between the control (93.5 percent) and test (93.5 percent) was less than 0.1 percentage points and is statistically significant.

Table 2. Original Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode

Mode	Control Interviews	Control Percent	Test Interviews	Test Percent	Test Minus Control	P-Value
All Modes	19,455	93.5 (0.3)	19,400	93.5 (0.3)	<0.1 (0.4)	0.98
Self-Response	13,284	53.7 (0.5)	13,131	52.9 (0.5)	-0.8 (0.6)	0.23
Internet	8,112	34.1 (0.4)	8,168	34.4 (0.4)	0.4 (0.6)	0.49
Mail	5,172	19.6 (0.3)	4,963	18.4 (0.3)	-1.2 (0.5)	0.01*
CATI	880	9.2 (0.4)	872	8.7 (0.4)	-0.4 (0.6)	0.44
CAPI	5,291	83.6 (0.6)	5,397	83.5 (0.7)	<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 showed a statistically significant difference was 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 high and low response areas. Table 3. shows the mail response rates for both treatments in high and low response areas.[26] 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

[26] Table C1 (including all modes) can be found in Appendix C.

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.

We noted 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 were no underlying response rate concerns that would impact those findings.

Table 3. Mail Response Rates by Designated High (HRA) and Low (LRA) Response Areas

	Control Interviews	Control Percent	Test Interviews	Test Percent	Test Minus Control	P-Value
HRA	2,224	21.5 (0.4)	2,082	20.0 (0.4)	-1.5 (0.6)	0.02*
LRA	2,948	14.1 (0.3)	2,881	13.8 (0.3)	-0.3 (0.4)	0.43
Difference	(N/A)	7.4 (0.4)	(N/A)	6.2 (0.5)	-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.

5.1.2 Unit Response Rates for the Content Follow-Up Interview

Table 4 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 4. Content Follow-Up Interview Unit Response Rates for Control and Test Treatments, Overall and by Mode of Original Interview

Original Interview Mode	Control Interviews	Control Percent	Test Interviews	Test Percent	Test Minus Control	P-Value
All Modes	7,903	45.7 (0.6)	7,867	44.8 (0.5)	-0.8 (0.8)	0.30
Internet	4,045	52.5 (0.7)	4,078	51.9 (0.6)	-0.6 (0.8)	0.49
Mail	2,197	44.2 (0.9)	2,202	46.4 (0.9)	2.1 (1.3)	0.11
CATI	399	51.5 (2.5)	369	48.9 (1.9)	-2.5 (2.9)	0.39
CAPI	1,262	37.7 (1.1)	1,218	34.8 (1.2)	-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 2016 ACS 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.*[27] The response distribution calculations can be found in Table 5. 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 is statistically significant.

[27] We were not able to conduct the demographic analysis shown in Table 5 by relationship status, race, or ethnicity because these topics were tested as part of the Content Test.

Table 5. Weighted and Unedited Response Distributions: Control Versus Test Treatment

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

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 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.

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 Table 6 and Table 7.

Table 6. Comparison of Average Household Size

	Control (n=17,608)	Test (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 testing based on a two-tailed t-test at the α=0.1 level.

Table 7. Comparison of Language of Response

Language of Response	Control Percent (n=17,694)	Test Percent (n=17,608)	Test Minus Control	P-value
English	96.2 (0.2)	96.1 (0.2)	<0.1 (0.3)	0.52
Spanish	2.6 (0.2)	2.7 (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 testing based on a two-tailed t-test at the α=0.1 level.

The 2016 ACS Content Test was available in two languages, English and Spanish, for all modes 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 were no detectable differences between control and test for *average household size* or *language of response*. There were 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 did exhibit similar demographic characteristics since none of the resulting findings is significant, which verified our assumption of demographic similarity between treatments.

5.2 Item Missing Data Rates

This section will answer the following research question: *Which question version has lower item missing data rates overall and by mode?*

Tables 8, 9, and 10 show the item missing data rates for the race and Hispanic origin questions. Item missing data rates include cases that were completely missing data and cases where the only data were uncodable. In Table 8, only the Hispanic origin question in the control treatment was compared with the combined question in the test treatment; in Table 9, only the race question in the control treatment was compared with the combined question in the test treatment. In Table 10, both the race and Hispanic origin questions in the control treatment (when a response was missing for both the Hispanic origin and race questions) were compared with the combined question in the test treatment.

Table 8. Item Missing Data Rates for the Control Hispanic Origin Question and the Test Combined Question, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	1.7 (0.1)	43,593	0.6 (0.1)	-1.1 (0.2)	<0.01*
Self-Response	31,484	2.6 (0.1)	31,228	0.6 (0.1)	-2.0 (0.2)	<0.01*
Internet	20,861	0.6 (0.1)	21,102	0.5 (0.1)	-0.1 (0.2)	1.00
Mail	10,623	6.9 (0.4)	10,126	0.9 (0.2)	-6.1 (0.5)	<0.01*
Interviewer Assisted	12,187	0.5 (0.2)	12,365	0.7 (0.2)	0.3 (0.3)	1.00
CATI	1,894	0.5 (0.3)	1,963	0.4 (0.2)	-0.2 (0.3)	1.00
CAPI	10,293	0.4 (0.2)	10,402	0.7 (0.2)	0.3 (0.3)	1.00

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 8 shows that when compared with just the Hispanic origin question in the control treatment, the test treatment has a lower overall item missing data rate. This was driven by the mail mode, as there were no significant differences between treatments for any other modes. This may be because non-Hispanic respondents may have chosen to skip the Hispanic origin question in the mail mode, while in other modes they were prompted for a response.

Table 9. Item Missing Data Rates for the Control Race Question and the Test Combined Question, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	1.4 (0.1)	43,593	0.6 (0.1)	-0.8 (0.2)	<0.01*
Self-Response	31,484	1.7 (0.1)	31,228	0.6 (0.1)	-1.1 (0.2)	<0.01*
Internet	20,861	1.1 (0.1)	21,102	0.5 (0.1)	-0.6 (0.2)	<0.01*
Mail	10,623	2.9 (0.3)	10,126	0.9 (0.2)	-2.0 (0.3)	<0.01*
Interviewer Assisted	12,187	1.1 (0.3)	12,365	0.7 (0.2)	-0.4 (0.3)	0.50
CATI	1,894	1.8 (0.8)	1,963	0.4 (0.2)	-1.4 (0.8)	0.22
CAPI	10,293	1.0 (0.3)	10,402	0.7 (0.2)	-0.3 (0.3)	0.50

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 9 shows that when compared with just the race question in the control treatment, the test treatment has a lower overall item missing data rate. The test treatment also had lower item missing data rates for self-response modes (mail and internet). These results suggest that the test treatment reduces the item missing data rate for the race question. This is most likely because many Hispanics choose not to answer the race question in the control treatment, which is not an issue in the combined question format (see Table 39).

Table 10. Item Missing Data Rates for both Race and Hispanic Origin Questions, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	0.4 (0.1)	43,593	0.6 (0.1)	0.2 (0.1)	0.30
Self-Response	31,484	0.5 (0.1)	31,228	0.6 (0.1)	0.1 (0.1)	1.00
Internet	20,861	0.5 (0.1)	21,102	0.5 (0.1)	<0.1 (0.2)	1.00
Mail	10,623	0.4 (0.1)	10,126	0.9 (0.2)	0.4 (0.2)	0.17
Interviewer Assisted	12,187	0.2 (0.2)	12,365	0.7 (0.2)	0.5 (0.2)	0.30
CATI	1,894	0.2 (0.1)	1,963	0.4 (0.2)	0.2 (0.2)	1.00
CAPI	10,293	0.3 (0.2)	10,402	0.7 (0.2)	0.5 (0.3)	0.30

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 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.

In Table 10, when comparing item missing data rates in the test treatment to both the Hispanic origin and race questions in the control treatment, there were no significant differences between the two treatments. This suggests that there were some respondents who chose not to answer the race and Hispanic origin questions, regardless of whether or not the question format was separate or combined. This also suggests that some respondents skip either the race or Hispanic origin question, but generally not both.

5.3 Response Proportions

5.3.1 Race Category Response Proportions

This section will answer the following research question: *How do the estimated proportions of persons in each of the six major race groups compare between control and test overall and by mode?*

The tables in this section show the response distributions for the six major race categories overall and by mode. These tables do not include the Hispanic or MENA responses. Hispanic and MENA tables will be presented in Section 5.3.2.

Table 11. Overall Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race

Race	Control Percent (n=43,671)	Test Percent (n=43,593)	Test Minus Control	Adjusted P-Value
White	74.0 (0.5)	67.7 (0.6)	-6.3 (0.7)	<0.01*
Black	12.8 (0.4)	12.0 (0.4)	-0.8 (0.5)	0.43
Asian	6.9 (0.3)	6.6 (0.3)	-0.3 (0.4)	0.48
AIAN	4.1 (0.3)	3.1 (0.2)	-1.0 (0.3)	0.02*
NHPI	0.4 (0.1)	0.3 (0.1)	-0.2 (0.1)	0.48
SOR	8.2 (0.4)	2.3 (0.2)	-5.9 (0.4)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Race refers to race alone or in combination, as the question allowed for multiple categories to be marked. The table does not include Hispanic and MENA respondents who did not provide a race. Columns will not sum to 100%. 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 11 shows that the overall response distributions were different between the control and test treatments for White, AIAN, and SOR respondents. In the control treatment 74.0 percent of respondents were White compared with 67.7 percent in the test treatment. Likewise, in the control treatment 8.2 percent of respondents were SOR compared with 2.3 percent in the test treatment. Previous research has shown that Hispanic respondents are more likely to choose White or SOR as their race when given separate questions for race and Hispanic (Compton et al., 2013). When Hispanics can choose just Hispanic as their identity, however, as in the test treatment, the proportions of White, AIAN, and SOR responses decrease. This may be due to differences in multiple-reporting by Hispanic respondents who also identify as White, AIAN, and SOR between the control and the test treatments (see Table 39 for the full race distribution of Hispanic respondents). There were no significant differences between the control and test treatments for Black, Asian, or NHPI.

Table 12. Mail Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race

Race	Control Percent (n=10,623)	Test Percent (n=10,126)	Test Minus Control	Adjusted P-Value
White	81.0 (0.7)	76.7 (0.9)	-4.4 (1.2)	<0.01*
Black	9.6 (0.5)	10.5 (0.5)	0.9 (0.7)	0.68
Asian	5.2 (0.5)	5.4 (0.5)	0.2 (0.7)	1.00
AIAN	2.5 (0.2)	2.1 (0.2)	-0.4 (0.3)	0.68
NHPI	0.2 (0.1)	0.2 (0.1)	<0.1 (0.1)	1.00
SOR	3.4 (0.4)	2.1 (0.2)	-1.2 (0.4)	0.02*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Race refers to race alone or in combination, as the question allowed for multiple categories to be marked. The table does not include Hispanic and MENA respondents who did not provide a race. Columns will not sum to 100%. 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.

According to Table 12, for those who responded by mail, the proportion of White responses was lower in the test treatment (76.7 percent) compared with the control treatment (81.0 percent). The proportion of SOR responses was also lower in the test treatment (2.1 percent) compared with the control treatment (3.4 percent). This is most likely a result of differences in Hispanic reporting between the control and test treatments. There were no significant differences between the two treatments for any of the other categories.

Table 13. Internet Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race

Race	Control Percent (n=20,861)	Test Percent (n=21,102)	Test Minus Control	Adjusted P-Value
White	80.8 (0.6)	77.5 (0.6)	-3.3 (0.8)	<0.01*
Black	7.2 (0.3)	6.9 (0.4)	-0.3 (0.5)	1.00
Asian	8.7 (0.5)	8.8 (0.4)	0.2 (0.5)	1.00
AIAN	3.4 (0.2)	2.9 (0.2)	-0.4 (0.3)	0.73
NHPI	0.5 (0.1)	0.3 (0.1)	-0.1 (0.1)	0.73
SOR	6.1 (0.3)	2.0 (0.2)	-4.1 (0.4)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Race refers to race alone or in combination, as the question allowed for multiple categories to be marked. The table does not include Hispanic and MENA respondents who did not provide a race. Columns will not sum to 100%. 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.

According to Table 13, the internet response distribution was also lower for both the White and SOR categories in the test treatment compared with the control treatment. As with the mail mode response distribution, there were no significant differences between the two treatments for the other categories.

Table 14. CATI Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race

Race	Control Percent (n=1,894)	Test Percent (n=1,963)	Test Minus Control	Adjusted P-Value
White	73.7 (2.3)	66.7 (2.2)	-7.1 (3.1)	0.12
Black	18.3 (2.0)	19.7 (2.1)	1.4 (2.9)	1.00
Asian	4.2 (1.2)	4.0 (1.3)	-0.2 (1.9)	1.00
AIAN	4.2 (0.8)	3.7 (0.7)	-0.5 (0.9)	1.00
NHPI	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	0.65
SOR	4.9 (0.9)	2.2 (0.5)	-2.7 (1.0)	0.04*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Race refers to race alone or in combination, as the question allowed for multiple categories to be marked. The table does not include Hispanic and MENA respondents who did not provide a race. Columns will not sum to 100%. 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.

In Table 14, the CATI response distribution showed a lower proportion of SOR responses for the test treatment (2.2 percent) compared with the control treatment (4.9 percent). There were no significant differences between treatments for any of the other categories. Also, CATI was the only mode where the proportion of White responses was not significantly lower for the test treatment.

Table 15. CAPI Response Distribution for Control and Test Treatments (Weighted and Unedited), by Race

Race	Control Percent (n=10,293)	Test Percent (n=10,402)	Test Minus Control	Adjusted P-Value
White	62.7 (1.2)	52.5 (1.4)	-10.2 (1.7)	<0.01*
Black	20.2 (1.0)	17.8 (1.1)	-2.5 (1.5)	0.29
Asian	6.0 (0.8)	4.9 (0.5)	-1.1 (0.8)	0.40
AIAN	5.8 (0.6)	3.9 (0.4)	-1.9 (0.8)	0.08*
NHPI	0.6 (0.3)	0.3 (0.1)	-0.3 (0.3)	0.40
SOR	13.4 (0.9)	2.7 (0.4)	-10.6 (0.9)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Race refers to race alone or in combination, as the question allowed for multiple categories to be marked. The table does not include Hispanic and MENA respondents who did not provide a race. Columns will not sum to 100%. 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 15 shows that for the CAPI response distribution, there were lower proportions of White, AIAN, and SOR responses in the test treatment compared with the control treatment. Again, this was most likely due to differences in Hispanic reporting between the control and test treatments. There were no significant differences between the test and control treatments for the Black, Asian, or NHPI categories.

5.3.2 Hispanic and MENA Reporting

The following research question aimed to examine the differences between Hispanic reporting in the control and test treatments by mode. Likewise, since the MENA category was added for the test treatment, it was also important to examine MENA reporting by mode.

This section will answer the following research question: *How do the estimates for the overall Hispanic origin and MENA populations compare between the control and test versions overall and by mode?*

Table 16. Percent Reporting Hispanic Origin in Test and Control Treatments, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	17.1 (0.5)	43,593	17.7 (0.5)	0.5 (0.7)	1.00
Self-Response	31,484	10.4 (0.4)	31,228	10.7 (0.3)	0.3 (0.5)	1.00
Internet	20,861	10.5 (0.4)	21,102	11.4 (0.4)	0.8 (0.6)	1.00
Mail	10,623	10.1 (0.6)	10,126	9.0 (0.6)	-1.1 (0.8)	1.00
Interviewer Assisted	12,187	27.5 (1.1)	12,365	28.1 (1.2)	0.7 (1.5)	1.00
CATI	1,894	13.1 (1.9)	1,963	12.0 (1.4)	-1.1 (2.4)	1.00
CAPI	10,293	28.6 (1.2)	10,402	29.4 (1.3)	0.8 (1.6)	1.00

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 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 16 shows that approximately 17 percent of respondents reported being Hispanic for both the combined format of the race or ethnicity question and the separate question format. There were no significant differences between the control or test treatment by any mode.

Table 17. Percent Reporting Middle Eastern or North African in Test and Control Treatments, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	1.1 (0.1)	43,593	1.4 (0.2)	0.3 (0.2)	1.00
Self-Response	31,484	1.1 (0.1)	31,228	1.3 (0.2)	0.2 (0.2)	1.00
Internet	20,861	1.4 (0.2)	21,102	1.6 (0.2)	0.2 (0.3)	1.00
Mail	10,623	0.5 (0.2)	10,126	0.8 (0.2)	0.3 (0.3)	1.00
Interviewer Assisted	12,187	1.1 (0.2)	12,365	1.5 (0.4)	0.4 (0.4)	1.00
CATI	1,894	0.5 (0.3)	1,963	0.7 (0.4)	0.2 (0.5)	1.00
CAPI	10,293	1.1 (0.3)	10,402	1.5 (0.4)	0.4 (0.5)	1.00

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: MENA responses in the control treatment were tabulated using write-in responses. Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance 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 17 shows the percentage of respondents reporting Middle Eastern or North African in both treatments. Note that the control treatment did not have a distinct MENA category, but the White category had two MENA examples (Lebanese and Egyptian) in the write-in area. Any MENA answers that were reported as either a checkbox (test treatment only) or a write-in (test and control treatments) were coded and tabulated for the analysis in Table 17. There were no statistical differences between the test and control treatments in the percent reporting MENA; approximately one percent of respondents identified as MENA in both treatments.

5.4 Benchmarks

This section will answer the following research question: *How do the estimated proportions of persons in each race or ethnic category for each treatment (control and test) compare with the proportions from the 2010 Census, the AQE, and the 2015 NCT?*

Table 18 shows data from the 2010 Census, the 2015 NCT, and the 2016 ACS Content Test. Note that, as discussed in Section 2.4.4, the 2010 AQE is not included in this table. The 2010 Census and the 2015 NCT included all states and the 2010 Census data were fully edited. The 2016 ACS Content Test did not include Alaska, Hawaii, or Puerto Rico, and the data were not fully edited.

In the 2016 ACS Content Test control treatment (separate question format), the 2010 Census, and the 2015 NCT separate question format (Panel 1), there was no MENA category; all MENA responses were included in the White category. In the test treatment of the 2016 ACS Content Test (combined question format) and the combined question format of the 2015 NCT (Panel 36), a MENA category was available. Therefore, for these columns in Table 18, the White category does not include MENA responses.

Table 18. Race and Ethnicity Percentages for the 2010 Census, the 2015 National Content Test (2015 NCT), and the American Community Survey Content Test (2016 ACS CT)

Race or Ethnicity	2010 Census	2015 NCT Separate	2015 NCT Combined	2016 ACS CT Separate	2016 ACS CT Combined
White (with MENA)	74.8	72.9 (0.7)	X	74.0 (0.5)	X
White (without MENA)	X	X	70.1 (0.8)	X	66.7 (0.6)
Hispanic	16.3	14.7 (0.4)	16.7 (0.4)	17.1 (0.5)	17.7 (0.5)
Black	13.6	13.3 (0.6)	12.9 (0.6)	12.8 (0.4)	12.0 (0.4)
Asian	5.6	6.0 (0.2)	5.9 (0.2)	6.9 (0.3)	6.6 (0.3)
AIAN	1.7	3.1 (0.1)	3.4 (0.1)	4.1 (0.3)	3.1 (0.2)
MENA	X	X	1.0 (<0.1)	X	1.4 (0.2)
NHPI	0.4	0.4 (<0.1)	0.4 (<0.1)	0.4 (0.1)	0.3 (0.1)
SOR	7.0	9.3 (0.2)	0.9 (<0.1)	8.2 (0.4)	2.3 (0.2)

Source: U.S. Census Bureau, 2010 Census; U.S. Census Bureau, 2015 National Content Test; U.S. Census Bureau, 2016 American Community Survey Content Test

Note: Race or ethnicity refers to race alone or in combination, as the question allowed for multiple categories to be marked. Columns may not sum to 100%. This table does not include data for Puerto Rico. An X indicates a category that was not available for that particular questionnaire. Standard errors are shown in parentheses. 2010 Census and 2015 NCT data were fully edited; 2016 ACS CT data were unedited. 2015 NCT Separate used data from panel 1 and 2015 NCT Combined used data from panel 36.

Table 18 shows the distributions of each of the major race and ethnic categories for the 2016 ACS Content Test control and test treatments, the 2010 Census, and the 2015 NCT separate and combined question versions. Recall that unlike the 2010 Census, the 2016 ACS Content Test and the 2015 NCT were not fully edited or imputed.

5.5 Response Error

This section will answer the following research question: *Using results from CFU, which question version results in estimates that are more reliable overall and by mode?*

This section compares the Gross Difference Rates (GDR) and Indexes of Inconsistency (IOI) between the control and test treatments. The results use data where the respondent was the same person for both interviews.

Table 19. Overall Difference in Gross Difference Rates (GDR) between Control and Test Treatments

Race or Ethnicity	Control GDR Percent (n=16,449)	Test GDR Percent (n=16,309)	Test Minus Control	Adjusted P-Value
White	5.4 (0.5)	3.0 (0.3)	-2.4 (0.5)	<0.01*
Hispanic	1.2 (0.2)	1.3 (0.3)	0.1 (0.3)	1.00
Black	0.9 (0.2)	0.8 (0.2)	-0.2 (0.3)	1.00
Asian	0.6 (0.1)	0.4 (0.1)	-0.1 (0.1)	1.00
AIAN	4.7 (0.3)	3.9 (0.4)	-0.8 (0.5)	0.69
MENA	0.5 (0.1)	0.2 (0.1)	-0.2 (0.1)	0.18
NHPI	0.3 (0.1)	0.2 (0.1)	-0.2 (0.1)	1.00
SOR	7.7 (0.5)	3.1 (0.3)	-4.6 (0.6)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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 19 shows that, overall, the GDR was significantly higher for White and SOR in the control treatment compared with the test treatment. In the control treatment, 5.4 percent of White responses were inconsistent between the original interview and the reinterview. In contrast, in the test treatment, the rate of inconsistent responses was 3.0 percent. The test treatment provided more reliable results of the White and SOR populations.

The following tables (Tables 20 – 23) show that these results were generally consistent across mode of original interview. The exceptions were that:

- For persons responding by mail originally, only White had a significant difference in GDR between control and test treatments;
- For persons responding by internet originally, only SOR had a significant difference in GDR between control and test treatments;
- For persons responding by CATI originally, no significant differences were identified between control and test GDR results; and
- For persons responding by CAPI originally, both White and SOR had significant differences between the control and test treatments.

There were no significant differences between treatments for any of the other race categories.

Table 20. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – Mail

Race or Ethnicity	Control GDR Percent (n=3,762)	Test GDR Percent (n=3,684)	Test Minus Control	Adjusted P-Value
White	3.0 (0.4)	1.3 (0.3)	-1.7 (0.5)	0.01*
Hispanic	0.9 (0.3)	1.0 (0.2)	0.2 (0.3)	1.00
Black	0.5 (0.2)	0.2 (0.1)	-0.3 (0.2)	0.43
Asian	0.5 (0.2)	0.5 (0.2)	<0.1 (0.3)	1.00
AIAN	4.2 (0.6)	3.5 (0.4)	-0.7 (0.8)	1.00
MENA	0.6 (0.2)	0.5 (0.3)	<0.1 (0.4)	1.00
NHPI	-	-	-	-
SOR	5.3 (0.7)	3.5 (0.5)	-1.8 (0.8)	0.17

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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. 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.

Table 21. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – Internet

Race or Ethnicity	Control GDR Percent (n=8,924)	Test GDR Percent (n=9,177)	Test Minus Control	Adjusted P-Value
White	2.7 (0.3)	2.2 (0.3)	-0.5 (0.4)	0.76
Hispanic	0.9 (0.2)	0.9 (0.1)	<0.1 (0.2)	1.00
Black	0.5 (0.1)	0.5 (0.2)	<0.1 (0.2)	1.00
Asian	0.6 (0.1)	0.5 (0.1)	-0.2 (0.2)	1.00
AIAN	3.4 (0.3)	3.0 (0.3)	-0.3 (0.5)	1.00
MENA	0.6 (0.1)	0.3 (0.1)	-0.3 (0.1)	0.35
NHPI	-	-	-	-
SOR	4.6 (0.4)	3.0 (0.3)	-1.6 (0.5)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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. 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.

Table 22. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – CATI

Race or Ethnicity	Control GDR Percent (n=713)	Test GDR Percent (n=653)	Test Minus Control	Adjusted P-Value
White	3.1 (0.8)	2.7 (1.0)	-0.4 (1.2)	1.00
Hispanic	1.2 (0.7)	0.6 (0.4)	-0.6 (0.8)	1.00
Black	0.6 (0.4)	0.5 (0.4)	-0.1 (0.5)	1.00
Asian	-	-	-	-
AIAN	5.4 (1.4)	3.4 (1.0)	-2.0 (1.7)	0.91
MENA	-	-	-	-
NHPI	-	-	-	-
SOR	-	-	-	-

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance 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. 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.

Table 23. Difference in Gross Difference Rates (GDR) between Control and Test Treatments – CAPI

Race or Ethnicity	Control GDR Percent (n=3,050)	Test GDR Percent (n=2,795)	Test Minus Control	Adjusted P-Value
White	9.9 (1.2)	4.7 (0.8)	-5.1 (1.4)	<0.01*
Hispanic	1.7 (0.5)	2.0 (0.6)	0.3 (0.8)	1.00
Black	1.7 (0.4)	1.4 (0.6)	-0.4 (0.7)	1.00
Asian	0.5 (0.1)	0.3 (0.1)	-0.2 (0.2)	1.00
AIAN	6.4 (0.9)	5.1 (0.9)	-1.3 (1.2)	1.00
MENA	0.3 (0.2)	<0.1 (<0.1)	-0.3 (0.2)	0.26
NHPI	-	-	-	-
SOR	12.7 (1.1)	2.9 (0.6)	-9.8 (1.3)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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. 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.

Table 24. Overall Difference in Index of Inconsistency (IOI) between Control and Test Treatments

Race or Ethnicity	Control IOI Percent (n=16,449)	Test IOI Percent (n=16,309)	Test Minus Control	Adjusted P-Value
White	15.8 (1.4)	7.3 (0.8)	-8.6 (1.6)	<0.01*
Hispanic	4.6 (0.8)	4.7 (0.9)	0.1 (1.2)	1.00
Black	4.7 (0.8)	4.3 (1.2)	-0.4 (1.4)	1.00
Asian	4.5 (0.7)	3.5 (0.6)	-1.0 (1.0)	1.00
AIAN	50.2 (3.9)	50.3 (3.5)	0.1 (5.4)	1.00
MENA	19.9 (3.1)	10.6 (3.3)	-9.3 (4.8)	0.31
NHPI	32.4 (15.1)	47.1 (15.5)	14.7 (22.8)	1.00
SOR	61.4 (3.1)	75.2 (4.6)	13.8 (5.4)	0.07*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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.

In Table 24, the results indicate that SOR and AIAN fall into the "high inconsistency" category, regardless of treatment. The results also indicate that NHPI falls into the "intermediate inconsistency" category. However, as Alaska and Hawaii were not included in the sample, the results for AIAN and NHPI should be interpreted with caution. White and SOR, the same race and ethnicity groups that showed differences in GDR, also had significantly different IOI metrics. The IOI metric for White is significantly higher in control cases. The IOI metric for persons identifying as SOR is significantly lower in control cases.

The following tables show the IOI results by mode.

Table 25. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – Mail

Race or Ethnicity	Control IOI Percent (n=3,762)	Test IOI Percent (n=3,684)	Test Minus Control	Adjusted P-Value
White	10.8 (1.7)	3.9 (0.9)	-6.9 (1.9)	<0.01*
Hispanic	5.4 (1.6)	7.0 (1.3)	1.7 (2.2)	1.00
Black	2.5 (0.9)	0.9 (0.3)	-1.6 (1.0)	0.63
Asian	6.5 (2.2)	5.8 (1.8)	-0.7 (2.9)	1.00
AIAN	61.8 (6.2)	55.4 (5.7)	-6.4 (8.4)	1.00
MENA	24.4 (10.6)	36.5 (20.9)	12.0 (23.4)	1.00
NHPI	-	-	-	-
SOR	76.5 (5.7)	83.8 (6.3)	7.3 (8.9)	1.00

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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. 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.

Table 26. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – Internet

Race or Ethnicity	Control IOI Percent (n=8,924)	Test IOI Percent (n=9,177)	Test Minus Control	Adjusted P-Value
White	10.5 (1.1)	6.8 (0.8)	-3.6 (1.3)	0.04*
Hispanic	6.0 (1.0)	5.2 (0.8)	-0.8 (1.2)	1.00
Black	4.1 (1.0)	4.3 (1.5)	0.2 (1.8)	1.00
Asian	4.5 (0.8)	3.1 (0.8)	-1.4 (1.2)	1.00
AIAN	45.2 (4.1)	46.9 (4.6)	1.7 (6.4)	1.00
MENA	16.9 (3.0)	11.6 (3.1)	-5.3 (4.8)	1.00
NHPI	-	-	-	-
SOR	54.9 (4.2)	83.9 (4.3)	29.0 (5.5)	<0.01*

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

Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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. 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.

Table 27. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – CATI

Race or Ethnicity	Control IOI Percent (n=713)	Test IOI Percent (n=653)	Test Minus Control	Adjusted P-Value
White	8.3 (2.1)	6.8 (2.5)	-1.5 (3.4)	1.00
Hispanic	6.8 (3.6)	3.4 (2.4)	-3.4 (4.4)	1.00
Black	1.8 (1.3)	1.6 (1.3)	-0.2 (1.7)	1.00
Asian	-	-	-	-
AIAN	45.6 (9.4)	33.0 (12.7)	-12.7 (18.3)	1.00
MENA	-	-	-	-
NHPI	-	-	-	-
SOR	-	-	-	-

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

Note: The results use data where the respondent was the same person for both interviews. Standard errors are shown in parentheses. Minor additive discrepancies are due to rounding. Significance 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. 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.

Table 28. Difference in Index of Inconsistency (IOI) between Control and Test Treatments – CAPI

Race or Ethnicity	Control IOI Percent (n=3,050)	Test IOI Percent (n=2,795)	Test Minus Control	Adjusted P-Value
White	22.8 (2.8)	9.6 (1.6)	-13.2 (3.1)	<0.01*
Hispanic	4.3 (1.2)	4.7 (1.5)	0.5 (2.0)	1.00
Black	6.0 (1.4)	5.8 (2.5)	-0.3 (2.7)	1.00
Asian	3.8 (1.4)	3.2 (1.6)	-0.6 (2.2)	1.00
AIAN	51.5 (7.8)	53.1 (6.1)	1.6 (10.2)	1.00
MENA	26.6 (9.7)	0.9 (1.1)	-25.7 (9.9)	0.05*
NHPI	-	-	-	-
SOR	64.1 (4.0)	63.1 (10.1)	-1.0 (11.2)	1.00

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: The results use data where the respondent was the same person for both interviews. MENA values in the control treatment were calculated using coded data from write-in responses. 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. 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.

Tables 25 to 28 show that the IOI metric for White showed the same trend across modes (higher in the control treatment for mail, internet, and CAPI). Table 26 shows that the IOI metric for SOR was higher in the test treatment for internet responses (consistent with the overall results). However, the assumption for parallel measures was not met for SOR in the internet mode, which indicates that this result is biased to some extent.

The above tables also show that the IOI metric was low across modes for Hispanic respondents and was not significantly different between the control and test treatments. This suggests that the reliability of Hispanic reporting was not impacted by treatment. There were no differences between the test and control treatments for any of the other race categories, except for MENA responses in the CAPI mode, which had a lower IOI for the test treatment (Table 28).

5.6 Results for Analysis Specific to Race and Hispanic Origin

5.6.1 Detailed Reporting

This section will answer the following research question: *How does detailed reporting for all race or ethnic categories (including Hispanic and MENA) proportionally differ overall and by mode?*

The tables in this section show the percentage of people reporting a detailed response overall and by mode for each race and ethnicity category. Two examples of detailed responses are when a respondent reported under the Black or African American category that he or she is Jamaican, or under the Hispanic category that he or she is Dominican. Detailed responses on the mail and internet forms came from write-ins or detailed checkboxes. Detailed responses reported in CATI and CAPI were recorded by the interviewer in response to open-ended questions about each group that listed examples of detailed categories (see Appendix B). Detailed responses did not have to correspond to the category under which they were reported in order to be counted. For

example, a respondent could have reported Jamaican under the NHPI category and still be counted as detailed Black or African American.

In order to clearly show MENA detailed reporting, MENA responses, regardless of where in the questionnaire they were provided by the respondent, are shown in a table separate from White responses. Therefore, MENA responses are shown in Table 35 and not included in Table 30, which shows White detailed responses.

Table 29. Percent of All Persons Reporting a Detailed Race or Ethnicity Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	75.0 (0.5)	43,593	80.6 (0.5)	5.6 (0.6)	<0.01*
Self-Response	31,484	68.9 (0.6)	31,228	79.3 (0.4)	10.4 (0.7)	<0.01*
Internet	20,861	81.6 (0.6)	21,102	93.3 (0.3)	11.7 (0.7)	<0.01*
Mail	10,623	41.4 (1.0)	10,126	46.4 (0.9)	5.0 (1.3)	<0.01*
Interviewer Assisted	12,256	84.3 (0.9)	12,296	82.5 (0.8)	-1.8 (1.1)	0.34
CATI	1,963	83.8 (1.9)	1,894	81.0 (2.0)	-2.7 (2.6)	0.34
CAPI	10,293	84.4 (1.0)	10,402	82.6 (0.9)	-1.7 (1.2)	0.34

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Question allowed multiple categories to be marked. Columns will not sum to 100%. 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 29 shows that overall, detailed reporting was higher for the test treatment compared with the control treatment (80.6 percent compared with 75.0 percent). Detailed responses were provided more often in the test treatment than in the control treatment for both mail and internet modes. In the test treatment, the internet version had both write-in spaces and detailed checkboxes, while the mail version only had write-in spaces. Therefore, there was more opportunity and more prompting for detailed responses in the internet mode. There were no significant differences between treatments for the CATI and CAPI modes.

Tables 30 through 37 show the percentage of respondents providing a detailed response for each race and ethnicity category. For example, 60.0 percent of respondents who indicated that they were White provided a detailed White response, such as German, in the control treatment.

Table 30. Percent of White Respondents Reporting a Detailed White Response (MENA Excluded), by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	29,077	60.0 (0.7)	25,316	74.1 (0.7)	14.1 (1.0)	<0.01*
Self-Response	22,661	60.2 (0.6)	20,673	76.4 (0.5)	16.2 (0.8)	<0.01*
Internet	15,167	74.6 (0.7)	14,073	91.3 (0.5)	16.7 (0.9)	<0.01*
Mail	7,494	29.5 (1.1)	6,600	41.3 (1.0)	11.8 (1.4)	<0.01*
Interviewer Assisted	6,416	59.5 (1.7)	4,643	69.1 (1.6)	9.5 (2.4)	<0.01*
CATI	1,255	72.0 (2.7)	1,026	77.0 (2.7)	5.0 (3.5)	0.15
CAPI	5,161	58.4 (1.8)	3,617	68.3 (1.8)	9.9 (2.6)	<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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 30 shows that there was a significantly higher proportion of White detailed reporting overall in the test treatment compared with the control treatment (74.1 percent compared with 60.0 percent). The same was true for all modes, with the exception of CATI responses.

Table 31. Percent of Hispanic Respondents Reporting a Detailed Hispanic Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	9,380	93.5 (0.7)	9,431	89.8 (0.8)	-3.7 (1.1)	<0.01*
Self-Response	5,071	89.7 (1.0)	5,000	85.6 (1.0)	-4.1 (1.4)	0.01*
Internet	3,449	91.3 (1.1)	3,609	92.6 (1.0)	1.3 (1.4)	0.35
Mail	1,622	86.1 (2.2)	1,391	64.6 (2.6)	-21.5 (3.2)	<0.01*
Interviewer Assisted	4,309	95.7 (0.8)	4,431	92.2 (1.0)	-3.5 (1.3)	0.03*
CATI	359	96.9 (1.4)	405	88.3 (3.2)	-8.6 (3.7)	0.05*
CAPI	3,950	95.7 (0.9)	4,026	92.3 (1.1)	-3.4 (1.4)	0.05*

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 31 shows that overall, there was a lower proportion of Hispanic detailed responses in the test treatment compared with the control treatment. Additionally, there was a lower proportion of detailed Hispanic responses for the mail mode in the test treatment (64.6 percent) compared with the control treatment (86.1 percent). This is likely because the respondents receiving the control version had access to detailed checkboxes in the Hispanic origin question, while the test treatment did not have detailed checkboxes.[28] There were no differences between the treatments for the internet mode, however, which had detailed checkboxes available for both control and test versions. There was also a lower proportion of Hispanic detailed responses in the test

[28] This design with multiple checkboxes and a write-in area for all groups was tested on paper in the 2015 NCT. Similar results were observed, indicating that the inclusion of checkboxes led to higher detailed reporting for Hispanics. The 2017 Census Test will further test the multi-checkbox with write-in design.

treatment compared with the control for the interviewer-assisted modes (CATI and CAPI). This is likely due to the absence of detailed checkboxes on the test treatments in the CATI and CAPI modes.

Table 32. Percent of Black or African American Respondents Reporting a Detailed Black or African American Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	7,456	72.2 (1.4)	7,286	81.6 (0.9)	9.4 (1.7)	<0.01*
Self-Response	4,248	61.2 (1.4)	4,135	76.2 (1.5)	15.0 (2.1)	<0.01*
Internet	2,447	73.5 (1.8)	2,359	98.0 (0.5)	24.5 (1.8)	<0.01*
Mail	1,801	41.1 (2.4)	1,776	42.2 (2.6)	1.2 (3.6)	0.82
Interviewer Assisted	3,208	78.9 (2.3)	3,151	85.2 (1.2)	6.3 (2.7)	0.06*
CATI	457	83.4 (4.9)	515	78.3 (3.4)	-5.1 (6.1)	0.82
CAPI	2,751	78.6 (2.4)	2,636	85.8 (1.3)	7.2 (2.9)	0.05*

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 32 shows that overall, there was a significantly higher proportion of Black detailed reporting in the test treatment compared with the control (81.6 percent compared with 72.2 percent). There was also higher Black detailed reporting in the test treatment in the internet and CAPI modes. There were no significant differences between the treatments in Black detailed reporting for the mail and CATI modes.

Table 33. Percent of Asian Respondents Reporting a Detailed Asian Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	2,826	97.5 (0.6)	2,814	95.0 (0.7)	-2.5 (1.0)	0.08*
Self-Response	2,271	97.0 (0.7)	2,308	94.9 (0.8)	-2.2 (1.2)	0.28
Internet	1,747	97.6 (0.8)	1,815	99.0 (0.4)	1.4 (0.9)	0.28
Mail	524	95.0 (1.9)	493	79.2 (3.3)	-15.8 (4.0)	<0.01*
Interviewer Assisted	555	98.4 (0.8)	506	95.3 (1.6)	-3.1 (1.7)	0.28
CATI	80	97.2 (3.1)	53	100.0 (<0.1)	2.8 (3.1)	0.35
CAPI	475	98.5 (0.8)	453	95.0 (1.8)	-3.5 (1.8)	0.28

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 33 shows that overall, there was a higher proportion of Asian detailed reporting in the control treatment compared with the test. This was driven mostly by the higher proportion of Asian detailed reporting in the control treatment for those who responded by mail. Like the control version of the Hispanic origin question, the race question for the control treatment included detailed Asian checkboxes, which were not included in the test version. There were no significant differences in Asian detailed reporting between the two treatments for any other mode.

Table 34. Percent of American Indian or Alaska Native (AIAN) Respondents Reporting a Detailed AIAN Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	1,683	67.6 (2.8)	1,371	61.4 (3.4)	-6.2 (4.2)	0.85
Self-Response	1,053	73.0 (2.4)	882	67.1 (2.4)	-5.9 (3.5)	0.64
Internet	781	75.9 (2.7)	641	70.4 (3.0)	-5.5 (4.1)	0.91
Mail	272	64.4 (5.3)	241	56.2 (4.8)	-8.2 (7.3)	1.00
Interviewer Assisted	630	63.2 (4.9)	489	55.5 (6.1)	-7.7 (7.9)	1.00
CATI	102	47.6 (8.1)	79	45.1 (9.2)	-2.5 (12.5)	1.00
CAPI	528	64.0 (5.1)	410	56.2 (6.4)	-7.8 (8.1)	1.00

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 34 shows that there were no significant differences in AIAN detailed reporting between the two treatments. Since Alaska was not included in the sample for the 2016 ACS Content Test, these results should be used with caution.

Table 35. Percent of Middle Eastern or North African (MENA) Respondents Reporting a Detailed MENA Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	399	95.3 (2.0)	506	91.6 (2.6)	-3.8 (3.3)	1.00
Self-Response	306	95.5 (2.6)	376	92.2 (2.5)	-3.3 (3.6)	1.00
Internet	269	97.2 (1.5)	295	94.8 (1.9)	-2.4 (2.2)	1.00
Mail	37	86.1 (12.4)	81	80.0 (10.5)	-6.0 (16.6)	1.00
Interviewer Assisted	93	95.0 (2.9)	130	90.7 (5.0)	-4.4 (6.0)	1.00
CATI	7	100.0 (<0.1)	7	100.0 (<0.1)	<0.1 (<0.1)	1.00
CAPI	86	94.9 (3.1)	123	90.3 (5.2)	-4.5 (6.2)	1.00

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 35 shows detailed Middle Eastern and North African reporting. For an explanation of the detailed groups that were included in this category, see Section 1.3. Please note that since the control treatment did not include a MENA checkbox, in order to be included in this analysis, respondents had to write-in a response that indicated MENA. There were no significant differences in MENA detailed reporting between the two treatments.

Table 36. Percent of Native Hawaiian or Other Pacific Islander (NHPI) Respondents Reporting a Detailed NHPI Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	151	82.4 (9.3)	118	83.2 (5.4)	0.9 (10.7)	1.00
Self-Response	105	92.9 (2.9)	83	75.9 (8.0)	-17.0 (8.8)	0.26
Internet	75	96.6 (2.0)	68	88.1 (6.1)	-8.5 (6.5)	0.77
Mail	30	73.2 (13.4)	15	20.1 (15.9)	-53.0 (21.6)	0.09*
Interviewer Assisted	46	72.1 (20.0)	35	92.3 (6.7)	20.2 (21.4)	1.00
CATI	-	-	-	-	-	-
CAPI	44	72.3 (20.1)	35	92.3 (6.7)	20.1 (21.6)	1.00

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. 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.

Table 36 shows that overall, there was no significant difference between the control and test treatments for NHPI. For the mail mode, there was a higher proportion of NHPI detailed reporting for the control treatment compared with the test treatment. This may be related to the availability of detailed checkboxes on the control treatment for the NHPI category, while there were none for the test treatment. There were no significant differences in NHPI detailed reporting between the two treatments for any other modes. Since Hawaii was not included in the sample for the 2016 ACS Content Test, these results should be used with caution.

Table 37. Percent of Some Other Race or Ethnicity (SOR) Respondents Reporting a Detailed SOR Response, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	4,428	19.5 (1.3)	986	65.2 (2.9)	45.7 (3.2)	<0.01*
Self-Response	2,256	30.8 (1.9)	674	54.8 (3.8)	24.0 (4.2)	<0.01*
Internet	1,748	30.2 (2.3)	439	52.4 (4.2)	22.2 (4.6)	<0.01*
Mail	508	33.0 (4.2)	235	60.1 (6.3)	27.1 (7.3)	<0.01*
Interviewer Assisted	2,172	12.4 (1.8)	312	76.8 (4.7)	64.3 (5.1)	<0.01*
CATI	124	13.2 (5.4)	45	68.2 (9.7)	55.0 (11.1)	<0.01*
CAPI	2,048	12.4 (1.8)	267	77.3 (4.8)	64.9 (5.2)	<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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 37 shows that there was a significantly higher proportion of SOR detailed reporting in the test treatment (65.2 percent) compared with the control treatment (19.5 percent). This result was the same across all modes. This may be because Hispanic respondents in the control treatment were choosing the SOR category and either writing in a generic response (i.e., Hispanic) or not providing a detailed response. Since there was a Hispanic category in the test treatment, however, fewer people responded as SOR and those who did were more likely to report non-

Hispanic responses and responses such as American, which are not classified under any of the OMB-defined race groups.[29]

5.6.2 Multiple-Response and Multiple-Race Reporting

This section will answer the following research question: *How do the proportions of multiple-response and multiple-race persons compare between control and test overall and by mode?*

Table 38 shows the proportion of respondents that had a multiple-response, which means they reported both a Hispanic origin and another race or ethnicity category. For example, a respondent could have reported that they were both Puerto Rican and White and be considered multiple-response.

Table 38. Multiple-Response Reporting for Control and Test Treatments, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	11.3 (0.5)	43,593	4.1 (0.2)	-7.2 (0.5)	<0.01*
Self-Response	31,484	7.0 (0.3)	31,228	3.2 (0.2)	-3.8 (0.3)	<0.01*
Internet	20,861	7.2 (0.4)	21,102	3.7 (0.2)	-3.5 (0.4)	<0.01*
Mail	10,623	6.7 (0.5)	10,126	2.1 (0.2)	-4.7 (0.5)	<0.01*
Interviewer Assisted	12,187	17.7 (1.1)	12,365	5.4 (0.5)	-12.3 (1.2)	<0.01*
CATI	1,894	9.5 (1.4)	1,963	3.0 (0.8)	-6.4 (1.6)	<0.01*
CAPI	10,293	18.4 (1.1)	10,402	5.6 (0.5)	-12.8 (1.2)	<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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Multiple-response reporting was higher in the control treatment than in the test treatment (11.3 percent compared with 4.1 percent) across all modes of data collection. These results suggest that a higher proportion of respondents identified as only Hispanic in the combined race and Hispanic origin question without also reporting a race. There was a high rate of multiple-response reporting in the CAPI control treatment (18.4 percent) compared with the CAPI test treatment (5.6 percent). This seems reflective of more Hispanic respondents being in the CAPI mode compared with other modes.

[29] In the control treatment, the three most common responses reported in the SOR write-in area were "Hispanic," "Mexican," and "Latino." In the test treatment, the three most common responses reported in the SOR write-in area were uncodable responses (such as proper names, "Don't know", "American," and "White").

Table 39. Percent of Hispanic Respondents Reporting a Race

Race	Control Percent (n=9380)	Test Percent (n=9431)	Test Minus Control	Adjusted P-Value
One Race	87.7 (0.8)	20.1 (1.0)	-67.6 (0.4)	<0.01*
White	51.7(1.8)	16.1 (1.0)	-35.6 (2.2)	<0.01*
Black	2.1 (0.5)	1.6 (0.3)	-0.6 (0.6)	1.00
Asian	0.6 (0.1)	0.7 (0.2)	0.1 (0.3)	1.00
AIAN	2.1 (0.5)	0.4 (0.1)	-1.7 (0.6)	0.01*
NHPI	0.1 (0.1)	0.1 (0.1)	<0.1 (0.1)	1.00
SOR	31.1 (1.6)	1.2 (0.3)	-29.9 (1.7)	<0.01*
Two or more races	7.2 (0.6)	3.1 (0.4)	-4.1 (0.7)	<0.01*
Missing Race	5.1 (0.5)	76.8 (1.1)	71.7 (1.3)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: MENA responses are tabulated under White. 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 39 shows the percentage of Hispanic respondents that reported each major race category. Of those Hispanic respondents that reported one race, significantly fewer reported White, AIAN, or SOR in the test treatment. There were no differences between treatments for Hispanic respondents reporting Black, Asian, or NHPI. There were also fewer Hispanics who reported two or more races in the test treatment. Additionally, significantly more Hispanic respondents only identified as Hispanic.

Table 40 shows the proportion of respondents that were multiple-race, which means they reported two or more races in the control treatment, or two or more races, not including Hispanic responses, in the test treatment. Two detailed responses within a major race category were not considered multiple-race. For instance, if someone reported that he or she was Korean and Chinese, that person would not be considered multiple-race.

Table 40. Multiple-Race Reporting for Control and Test Treatments, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	7.3 (0.3)	43,593	5.9 (0.2)	-1.4 (0.4)	<0.01*
Self-Response	31,484	6.6 (0.3)	31,228	5.8 (0.2)	-0.8 (0.3)	0.05*
Internet	20,861	7.6 (0.3)	21,102	6.4 (0.3)	-1.1 (0.4)	0.03*
Mail	10,623	4.5 (0.4)	10,126	4.3 (0.4)	-0.2 (0.5)	0.88
Interviewer Assisted	12,187	8.4 (0.6)	12,365	6.2 (0.5)	-2.3 (0.9)	0.04*
CATI	1,894	6.1 (0.9)	1,963	5.2 (0.8)	-0.8 (1.1)	0.88
CAPI	10,293	8.6 (0.6)	10,402	6.2 (0.6)	-2.4 (0.9)	0.04*

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 40 shows that a higher proportion of respondents in the control treatment reported being multiple-race (7.3 percent) compared with respondents in the test treatment (5.9 percent). There were no significant differences in multiple-race reporting among the mail and CATI modes. In

the internet mode, 7.6 percent of respondents in the control treatment reported as multiple-race, compared with 6.4 percent in the test treatment. In CAPI mode, 8.6 percent reported multiple-race in the control treatment, compared with 6.2 percent in the test treatment. The lower percentage of multiple-race reporting in the test could be due to fewer Hispanics checking SOR as well as another race category.

5.6.3 Ancestry Information

This section will answer the following research questions: *How does the information obtained from the race and Hispanic origin question write-ins in the control version compare with the information obtained from the ancestry question overall and by mode? How does information obtained from the race and Hispanic origin question write-ins in the test version compare with information obtained from the ancestry question overall and by mode?*

Both the control and test versions of the survey included an ancestry question that was asked several questions after the race and Hispanic origin questions (see Figure 3). The question asked, "What is this person's ancestry or ethnic origin?" Respondents were provided with two write-in lines and up to two ancestries were tabulated. The ancestry question was asked in the decennial census from 1980 to 2000, and since then has been asked on the American Community Survey. It is currently the only Census Bureau data source for detailed ethnic groups for the Black and White populations.

Figure 3. Ancestry Question

Table 41 shows the percentage of respondents in each treatment who reported race or ethnicity responses that were consistent with their ancestry response. In order to be included in the table, a respondent must have provided a race or ethnicity as well as an ancestry response. In order to be considered "consistent," their responses to the race and Hispanic origin question(s) must have exactly matched their response to the ancestry question, although they did not have to be in the same order or spelled exactly the same. For example, a response of Mexican and German in race or ethnicity was considered consistent with a response of German and Mexican in ancestry. A response of Hispanic (with no more detail) and German in race or ethnicity was not considered consistent with a response of German and Mexican in ancestry. The purpose of measuring consistency in this way was to determine whether respondents provided the same level of information in the race and ethnicity questions as they did in the ancestry question.

Table 41. Percent of Respondents with Consistent Race or Ethnicity and Ancestry Responses for Control and Test Treatments, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	36,231	72.5 (0.5)	35,599	74.0 (0.5)	1.5 (0.7)	0.09*
Self-Response	25,229	65.8 (0.5)	24,537	68.0 (0.5)	2.2 (0.7)	0.01*
Internet	16,990	73.3 (0.5)	16,676	73.8 (0.6)	0.5 (0.8)	1.00
Mail	8,239	48.8 (1.1)	7,861	54.1 (1.2)	5.3 (1.4)	<0.01*
Interviewer Assisted	11,002	81.9 (0.9)	11,062	82.2 (0.8)	0.4 (1.3)	1.00
CATI	1,659	83.6 (1.5)	1,703	77.0 (1.7)	-6.5 (2.3)	0.03*
CAPI	9,343	81.7 (0.9)	9,359	82.7 (0.9)	0.9 (1.3)	1.00

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

Table 41 shows that overall, respondents who answered the test version of the race and ethnicity questions reported more consistently with ancestry than those who answered the control version of the question(s). The test version resulted in 74.0 percent of respondents providing consistent responses, compared with 72.5 percent of respondents in the control treatment.

Respondents who answered by mail were more likely to have a race or ethnicity that was consistent with ancestry in the test treatment (54.1 percent) compared with those in the control treatment (48.8 percent). Among CATI respondents the trend was reversed; 77.0 percent of CATI respondents who received the test treatment reported a race or ethnicity consistent with ancestry, compared with 83.6 percent who received the control treatment.

This section will answer the following research question: *Does the item missing data rate for Ancestry differ between control and test versions overall and by mode?*

Table 42. Ancestry Item Missing Data Rates for Control and Test Treatments, by Mode

Mode	Control Sample Size	Control Percent	Test Sample Size	Test Percent	Test Minus Control	Adjusted P-Value
All Modes	43,671	13.2 (0.4)	43,593	15.7 (0.4)	2.5 (0.5)	<0.01*
Self-Response	31,484	15.2 (0.4)	31,228	17.4 (0.4)	2.3 (0.6)	<0.01*
Internet	20,861	14.1 (0.5)	21,102	17.1 (0.5)	3.0 (0.7)	<0.01*
Mail	10,623	17.6 (0.6)	10,126	18.3 (0.5)	0.7 (0.8)	0.79
Interviewer Assisted	12,187	10.3 (0.7)	12,365	13.1 (0.8)	2.8 (1.0)	0.02*
CATI	1,894	11.8 (1.6)	1,963	12.6 (1.6)	0.8 (2.5)	0.79
CAPI	10,293	10.1 (0.7)	10,402	13.1 (0.8)	3.0 (1.1)	0.02*

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. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

A higher proportion of people were missing ancestry responses in the test treatment than in the control treatments (Table 42). Of those that received the control treatment, 13.2 percent did not respond to the ancestry question, compared with 15.7 percent of those who received the test

treatment.[30] This was perhaps due to the design of the race and ethnicity question in the test treatment that allowed respondents to report any and all identities within the race question; by the time respondents got to the ancestry question, some may have believed they had already answered about their identity and not provided any further ancestry information.

Missing data rates were higher among test cases for the internet mode (14.1 percent in the control treatment and 17.1 percent in the test treatment) and in the CAPI mode (10.1 percent in control and 13.1 percent in the test). The mail and CATI modes did not have statistically different ancestry missing data rates between control and test treatments.

5.6.4 Demographic Information

This section will answer the following research question: *How do the estimates of race and Hispanic origin crossed by socioeconomic and demographic characteristics differ overall and by mode?*

Tables 43-46 show race and ethnicity reporting for respondents with selected characteristics from the ACS. These included age, sex, educational attainment, and housing tenure. Characteristics such as income and occupation were not included in these tables because these topics were being tested as part of the 2016 ACS Content Test.

Respondents who did not report a specific characteristic were not included in the table showing that characteristic.

[30] The percent of the population who did not report an ancestry in the 2015 American Community Survey was 14.4 percent.

Table 43. Race and Ethnicity Reporting in Test and Control Treatments, by Age

Category	Control Percent (n=43,325)	Test Percent (n=43,236)	Test Minus Control	Adjusted P-Value
Under 18	(n=9,774)	(n=9,767)		
White	70.9 (1.0)	61.7 (1.2)	-9.2 (1.4)	<0.01*
Hispanic	24.4 (1.0)	25.4 (1.1)	0.9 (1.4)	1.00
Black	15.5 (0.8)	14.8 (0.9)	-0.7 (1.2)	1.00
Asian	7.9 (0.6)	7.2 (0.5)	-0.7 (0.7)	1.00
AIAN	5.5 (0.6)	3.1 (0.3)	-2.4 (0.7)	<0.01*
MENA	1.2 (0.2)	1.4 (0.2)	0.2 (0.4)	1.00
NHPI	0.9 (0.3)	0.5 (0.1)	-0.4 (0.4)	1.00
SOR	10.8 (0.7)	2.4 (0.3)	-8.4 (0.8)	<0.01*
Age 18-34	(n=9,074)	(n=9,114)		
White	69.5 (0.9)	61.6 (1.1)	-8.0 (1.3)	<0.01*
Hispanic	20.7 (0.8)	21.1 (0.8)	0.4 (1.2)	1.00
Black	14.0 (0.5)	12.7 (0.7)	-1.3 (0.9)	0.83
Asian	8.4 (0.6)	8.1 (0.6)	-0.3 (0.8)	1.00
AIAN	4.4 (0.4)	3.6 (0.3)	-0.8 (0.6)	0.83
MENA	1.3 (0.2)	2.0 (0.4)	0.6 (0.4)	0.83
NHPI	0.4 (0.1)	0.4 (0.1)	<0.1 (0.1)	1.00
SOR	10.3 (0.5)	1.9 (0.2)	-8.4 (0.5)	<0.01*
Age 35-64	(n=17,098)	(n=16,906)		
White	75.0 (0.5)	67.9 (0.7)	-7.1 (0.8)	<0.01*
Hispanic	14.9 (0.5)	15.3 (0.5)	0.4 (0.6)	1.00
Black	11.9 (0.4)	11.0 (0.5)	-0.8 (0.6)	0.74
Asian	6.6 (0.4)	6.3 (0.3)	-0.3 (0.4)	1.00
AIAN	3.6 (0.3)	3.3 (0.2)	-0.3 (0.4)	1.00
MENA	1.0 (0.1)	1.3 (0.2)	0.3 (0.2)	1.00
NHPI	0.4 (0.1)	0.2 (<0.1)	-0.2 (0.1)	0.41
SOR	7.4 (0.4)	2.5 (0.2)	-4.9 (0.4)	<0.01*
Age 65+	(n=7,379)	(n=7,449)		
White	83.7 (0.6)	78.8 (0.7)	-4.9 (0.8)	<0.01*
Hispanic	7.2 (0.4)	8.0 (0.5)	0.8 (0.7)	1.00
Black	8.9 (0.5)	8.8 (0.5)	-0.1 (0.6)	1.00
Asian	4.0 (0.4)	4.5 (0.4)	0.6 (0.6)	1.00
AIAN	2.9 (0.3)	2.3 (0.3)	-0.6 (0.4)	1.00
MENA	0.9 (0.3)	0.9 (0.2)	<0.1 (0.3)	1.00
NHPI	0.1 (<0.1)	0.1 (<0.1)	<0.1 (0.1)	1.00
SOR	3.4 (0.3)	1.8 (0.3)	-1.6 (0.3)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Question allowed multiple categories to be marked. Columns will not sum to 100%. 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 43 shows that those who were under 18 years of age were less likely to report being White, AIAN, or SOR in the test treatment than in the control treatment. For the other age groups, however, the differences were only in White and SOR responses.

Table 44. Race and Ethnicity Reporting in Test and Control Treatments, by Sex

Category	Control Percent (n=43,456)	Test Percent (n=43,374)	Test Minus Control	Adjusted P-Value
Male	(n=21,024)	(n=20,885)		
White	74.4 (0.6)	66.9 (0.8)	-7.5 (1.0)	<0.01*
Hispanic	17.6 (0.6)	17.6 (0.6)	<0.1 (0.8)	1.00
Black	12.2 (0.4)	11.5 (0.5)	-0.7 (0.6)	0.94
Asian	6.8 (0.4)	6.7 (0.3)	-0.1 (0.4)	1.00
AIAN	4.0 (0.3)	3.1 (0.2)	-0.9 (0.4)	0.09*
MENA	1.3 (0.2)	1.6 (0.2)	0.4 (0.3)	0.94
NHPI	0.5 (0.1)	0.3 (0.1)	-0.2 (0.1)	0.94
SOR	8.6 (0.4)	2.4 (0.2)	-6.2 (0.5)	<0.01*
Female	(n=22,432)	(n=22,489)		
White	73.7 (0.6)	66.6 (0.6)	-7.0 (0.7)	<0.01*
Hispanic	16.8 (0.5)	17.7 (0.6)	1.0 (0.7)	0.74
Black	13.3 (0.5)	12.4 (0.5)	-0.9 (0.6)	0.69
Asian	7.0 (0.4)	6.5 (0.3)	-0.4 (0.5)	0.74
AIAN	4.2 (0.3)	3.2 (0.2)	-1.0 (0.4)	0.08
MENA	1.0 (0.1)	1.2 (0.2)	0.2 (0.2)	0.74
NHPI	0.5 (0.1)	0.3 (0.1)	-0.2 (0.2)	0.74
SOR	7.9 (0.4)	2.2 (0.2)	-5.7 (0.4)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Question allowed multiple categories to be marked. Columns will not sum to 100%. 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 44 shows that both males and females who received the test treatment were less likely to report being White, AIAN, or SOR than males and females who received the control treatment.

Table 45. Race and Ethnicity Reporting in Test and Control Treatments, by Educational Attainment (Persons 25 and Older)

Category	Control Percent (n=27,801)	Test Percent (n=27,482)	Test Minus Control	Adjusted P-Value
High School Degree or Less	(n=11,141)	(n=10,967)		
White	71.7 (0.7)	62.5 (0.9)	-9.1 (1.2)	<0.01*
Hispanic	22.2 (0.8)	22.2 (0.8)	<0.1 (1.2)	1.00
Black	13.2 (0.5)	12.2 (0.5)	-0.9 (0.6)	0.75
Asian	4.2 (0.4)	4.2 (0.4)	<0.1 (0.6)	1.00
AIAN	4.1 (0.3)	3.4 (0.3)	-0.7 (0.5)	0.75
MENA	0.9 (0.2)	0.6 (0.1)	-0.3 (0.2)	0.75
NHPI	0.4 (0.1)	0.1 (<0.1)	-0.3 (0.1)	0.04*
SOR	10.7 (0.6)	2.5 (0.3)	-8.3 (0.6)	<0.01*
Some College or Associate's Degree	(n=8,086)	(n=8,048)		
White	79.6 (0.7)	73.4 (0.8)	-6.1 (1.0)	<0.01*
Hispanic	11.4 (0.6)	11.7 (0.5)	0.3 (0.7)	1.00
Black	12.8 (0.6)	12.3 (0.6)	-0.5 (0.8)	1.00
Asian	3.6 (0.3)	4.2 (0.4)	0.6 (0.5)	1.00
AIAN	4.7 (0.3)	3.8 (0.3)	-0.9 (0.5)	0.34
MENA	0.5 (0.1)	1.0 (0.2)	0.5 (0.3)	0.56
NHPI	0.4 (0.1)	0.2 (0.1)	-0.2 (0.2)	1.00
SOR	5.6 (0.4)	2.1 (0.2)	-3.6 (0.4)	<0.01*
Bachelor's Degree or Higher	(n=8,574)	(n=8,467)		
White	80.4 (0.7)	77.3 (0.7)	-3.1 (1.0)	0.01*
Hispanic	6.8 (0.5)	7.0 (0.3)	0.2 (0.6)	1.00
Black	7.6 (0.5)	6.6 (0.4)	-1.0 (0.7)	0.65
Asian	11.2 (0.6)	10.1 (0.4)	-1.1 (0.7)	0.56
AIAN	2.3 (0.2)	2.2 (0.2)	-0.1 (0.3)	1.00
MENA	1.7 (0.2)	2.0 (0.3)	0.3 (0.4)	1.00
NHPI	0.2 (0.1)	0.2 (0.1)	<0.1 (0.1)	1.00
SOR	3.8 (0.3)	1.8 (0.2)	-2.0 (0.4)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
Note: Question allowed multiple categories to be marked. Columns will not sum to 100%. 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 45 shows that respondents with a high school degree or less education were less likely to report White, NHPI, or SOR in the test treatment compared with their counterparts in the control treatment. Respondents with some college or an associate's degree were less likely to report White or SOR in the test treatment compared with those in the control treatment. Likewise, respondents with a bachelor's degree or higher education were also less likely to report White or SOR in the test treatment compared with the control treatment. Since Hawaii was not included in the sample, the result for NHPI should be interpreted with caution.

Table 46. Race and Ethnicity Reporting in Test and Control Treatments, by Housing Tenure

Category	Control Percent (n=17,232)†	Test Percent (n=17,183)†	Test Minus Control	Adjusted P-Value
Owner-occupied housing units	(n=10,818)	(n=10,857)		
White	83.8 (0.4)	78.5 (0.5)	-5.4 (0.6)	<0.01*
Hispanic	8.8 (0.3)	9.6 (0.4)	0.8 (0.5)	0.37
Black	8.1 (0.3)	7.6 (0.3)	-0.5 (0.5)	0.77
Asian	5.1 (0.3)	4.7 (0.2)	-0.4 (0.4)	0.77
AIAN	3.2 (0.2)	2.7 (0.2)	-0.5 (0.3)	0.32
MENA	0.8 (0.1)	1.3 (0.2)	0.5 (0.2)	0.23
NHPI	0.1 (<0.1)	0.2 (0.1)	0.1 (0.1)	0.77
SOR	4.1 (0.3)	2.0 (0.2)	-2.1 (0.3)	<0.01*
Renter-occupied housing units	(n=6,414)	(n=6,326)		
White	65.3 (0.7)	58.6 (1.0)	-6.6 (1.2)	<0.01*
Hispanic	19.4 (0.7)	19.0 (0.7)	-0.4 (0.9)	1.00
Black	19.4 (0.7)	18.8 (0.8)	-0.6 (0.9)	1.00
Asian	6.9 (0.5)	6.2 (0.4)	-0.7 (0.7)	1.00
AIAN	5.0 (0.4)	3.7 (0.4)	-1.3 (0.7)	0.32
MENA	1.3 (0.2)	1.2 (0.2)	-0.1 (0.3)	1.00
NHPI	0.5 (0.2)	0.3 (0.1)	-0.2 (0.2)	0.84
SOR	10.7 (0.6)	2.9 (0.3)	-7.8 (0.7)	<0.01*

Source: U.S. Census Bureau, 2016 American Community Survey Content Test
†Housing units are classified based on the race of the householder. A small number of housing units did not have a householder identified. Thus, this universe is slightly smaller than the universe of housing units in Table 5 in Section 5.1.3.
Note: Question allowed multiple categories to be marked. Columns will not sum to 100%. 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 46 shows that for both owner- and renter-occupied housing units, respondents in the test treatment were less likely to report White and SOR than those in the control treatment.

Overall, the differences between the control and test treatments for the demographic and socioeconomic groups followed the same trend as the differences observed in the total population. There was lower reporting in the White and SOR categories, likely due to many people with Hispanic origin reporting only their Hispanic origin in the test treatment.

6 CONCLUSIONS

The 2016 ACS Content Test results for race and Hispanic origin confirm the conclusions made by earlier census test analyses, namely the 2010 AQE and the 2015 NCT. Race reporting for three of the five OMB-designated race groups (Black, Asian, and NHPI) did not differ between the two treatments. For the remaining two groups, White and AIAN, there were lower proportions when a combined race and Hispanic question was used. This appears to be due mainly to Hispanic respondents identifying with *only* the Hispanic category on the combined race and ethnicity question. As seen in the 2010 AQE and the 2015 NCT, this more accurately represents how Hispanic respondents self-identify. Furthermore, the level of White reporting in

the combined question format reflects the typical level of reporting for the non-Hispanic White population when there are separate Hispanic origin and race questions.

Results for the SOR category also confirm the findings from the 2010 AQE and the 2015 NCT. The SOR category was lower when a combined question was presented than when there were two separate questions. This finding was consistent across all modes. In effect, the SOR category was reduced to being the residual category it was intended to be, rather than being one of the largest groups.

Even though differences in Hispanic reporting contributed to lower proportions of White and SOR responses in the test treatment, the proportion reporting that they were Hispanic did not differ by treatment. Approximately 17 percent reported Hispanic origin in both the separate and combined questions. The level of Hispanic reporting did not differ by treatment for any of the modes.

The proportion of the population reporting that they were MENA did not differ between treatments, despite the fact that the test treatment had a distinct MENA category whereas the control treatment did not. There were no significant differences in MENA reporting between the treatments for any mode.

Item missing data rates represent respondents' understanding and willingness to answer the questions. Compared with the separate race question, the combined race and Hispanic origin question had higher data quality overall, as evidenced by lower item missing data rates. When comparing item missing data rates for the combined question with those for the separate race and Hispanic origin questions in the control treatment, no significant differences were detected overall. These findings line up with those of the 2010 AQE, which also suggested that combining the race and Hispanic origin questions into one question would reduce rates of item nonresponse (Compton et al., 2013).

The reliability measures also generally favored the test treatment. Specifically, the GDR and IOI measures from the response error reinterview indicated that the test and control treatments were not significantly different for Hispanic, Black, Asian, AIAN, and NHPI reporting and that the test treatment was more reliable for White.

Multiple-response reporting, when someone reported both a Hispanic origin and a race, was lower in the combined question format. This indicates that when presented with the combined question, fewer Hispanics report a race, but those who want to report both a Hispanic origin and a race continue to do so. As mentioned previously, this is more in line with how people of Hispanic origin tend to self-identify, as demonstrated through reinterview research with the 2010 AQE and the 2015 NCT, as well as extensive qualitative research with the 2010 AQE. Similarly, multiple-race reporting was also significantly higher in the control treatment compared with the test treatment overall and for the internet and CAPI modes.

Overall, detailed reporting was higher in the test treatment than in the control treatment. Detailed reporting was also higher in the test treatment for the mail and internet modes. Overall, there was higher detailed reporting in the test treatment for White, Black, and SOR and there were no

significant differences in detailed reporting for AIAN, MENA, and NHPI. However, detailed reporting was lower in the test treatment for Hispanic and Asian. This last result appears to be driven by the mail mode, as the mail version of the control treatment had detailed checkboxes for both Hispanic origin and Asian, but the test treatment did not.

The 2010 AQE also found lower detailed reporting among Hispanic and Asian respondents who were asked a combined question with no detailed checkboxes. This led to a new design for the combined question for the 2015 NCT in which the combined question offered multiple detailed checkboxes under each major OMB category. This 2015 NCT treatment resulted in increased detailed reporting for Hispanic and Asian respondents compared with the combined question with only write-in areas, which was the version tested in the 2016 ACS Content Test mail mode (Mathews et al., 2017). In the 2016 ACS Content Test internet mode, however, detailed checkboxes were available and resulted in similar levels of detailed reporting for Hispanic and Asian respondents in both treatments, confirming what was found with the 2015 NCT.

In addition to the results discussed above, there were also several analyses that were unique to the 2016 ACS Content Test. The 2016 ACS Content Test provided the opportunity to test the combined question and distinct MENA category in CATI and CAPI modes, which had only been tested in internet and mail modes in the 2015 NCT. The race and Hispanic origin response distributions for the control and test treatments by mode show that CATI and CAPI generally followed the trend of the overall response distribution. The same held true for the reliability and detailed reporting results. Also, results from other analyses in this report do not suggest any problems with collecting data in a combined question format in the CATI and CAPI modes.

The major finding from the socioeconomic and demographic analyses was that White and SOR reporting were lower in the test treatment for every age, sex, education, and tenure group. This result generally reflected the results for the total population in that White and SOR reporting was lower in the test treatment while reporting among the other groups was largely unaffected by treatment. The exceptions were that AIAN reporting was lower in the test treatment for those under age 18, and NHPI reporting was lower in the test treatment for those with a high school degree or less education. Since Alaska and Hawaii were not included in the sample, there was underrepresentation of these two groups in the sample. Therefore, the results for AIAN and NHPI should be interpreted with caution.

Race and Hispanic origin was compared with data from the ancestry question, and among respondents who were presented with the combined question format, responses were consistent with ancestry 74.0 percent of the time. The consistency was slightly lower for respondents who answered the separate race and Hispanic origin questions; they were consistent with ancestry 72.5 percent of the time. The item missing data rate for ancestry was higher in the test treatment than in the control treatment. Future research should examine whether people are less inclined to respond to the ancestry question when they have provided more detail about race and ethnicity, as well as analyze other responses inconsistencies to determine how those people were reporting.

Taken together, these results lead to the conclusion that the test treatment yields more accurate and reliable race and ethnicity reporting. OMB expects to make a decision about whether or not to revise the standards on question format and make MENA a minimum required category by the

end of 2017. Should the standards change, the results of this study indicate that the combined question with a MENA category could be successfully implemented in the American Community Survey.

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, Application Development and Services Division, Decennial Information Technology Division, Decennial Statistical Studies Division, Field Division, National Processing Center, Population Division, and Social, Economic, and Housing Statistics Division.
- Representatives from other agencies in the Federal statistical system serving on 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 formation and review of the Race and Hispanic Origin report:

- Karen Battle, Michael Bentley, Dorothy Barth, Joseph Brunn, Asaph Young Chun, Megan Darnell, Nicholas Jones, Colleen Keating, Rachel Marks, Jennifer Ortman, Eli Poehler, Roberto Ramirez, Jennifer Reichert, Andrew Roberts, Nicole Scanniello, Shabnam Shenasi Azari, Anthony Tersine, and Laquitta Walker.

8 REFERENCES

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

Buchanan, A., Marks, R., & Álvarez Figueroa, M. (2016). *2015 Forum on Ethnic Groups from the Middle East and North Africa: Meeting Summary and Main Findings*. U.S. Census Bureau, Washington, D.C. Retrieved June 15, 2017, from https://www.census.gov/library/working-papers/2015/demo/2015-MENA-Experts.html

Compton, E., Bentley, M., Ennis, S., & Rastogi, S. (2013). *2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment.* U.S. Census Bureau, Washington, D.C. Retrieved June 26, 2015, from http://www.census.gov/2010census/pdf/2010_Census_Race_HO_AQE.pdf

Dusch, G. and Meier, F. (2012). *2010 Census Content Reinterview Survey Evaluation Report*, U.S. Census Bureau, Washington, D.C. Retrieved June 8, 2017 from https://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, Washington, D.C. 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 January 31, 2017 from https://www.jstor.org/stable/4615733?seq=1#page_scan_tab_contents

Mathews, K., Phelan, J., Jones, N., Konya, S., Marks, R., Pratt, B., & Bentley, M. (2017). *2020 Research and Testing: 2015 National Content Test Race and Ethnicity Analysis Report.* U.S. Census Bureau, Washington, D.C. Retrieved June 8, 2017 from https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/final-analysis/2015nct-race-ethnicity-analysis.html

Myers, M., Trejo, Y.G., Lykke, L., Holliday, N., & Fernandez, L. (2015). *Cognitive and Usability Results from Spanish Pre-Testing of the 2015 National Content Test.* U.S. Census Bureau, Washington, D.C. Center for Survey Management, Research and Methodology Directorate.

Pratt, B., Hixson, L., & Jones, N. (2015). *Measuring Race and Ethnicity Across the Decades.* U.S. Census Bureau, Washington, D.C. Retrieved August 31, 2017 from https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html

Reiser, C., Barron, S., Bentley, M., Hill, J. M., Meier, A., & Pape, T. (2013). *2012 National Census Test Contact Strategy Results.* U.S. Census Bureau, Washington, D.C.

Sha, M., Schoua-Glusberg, A., Gerber, E., Kenward, K., Engel, J., Glusberg, D., & Corado, C. (2016). *Census 2020 Multilingual Research: Project 2 English and Spanish Cognitive and Usability Testing, Final Report.* RTI Project Number 0212349.009.

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

U.S. Census Bureau. (2014). *American Community Survey Design and Methodology (January 2014).* Retrieved February 1, 2017 from 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 June 8, 2017 from https://www.census.gov/research/data/planning_database/2015/

U.S. Office of Management and Budget. (1997). *1997 Revisions to the Standards for the Classification of Federal Data.*

Appendix A: Internet Versions of the Control and Test Questions

Figure A-1. Internet Version of the Control Treatment Hispanic Origin Question

Figure A-2. Internet Version of the Control Treatment Race Question

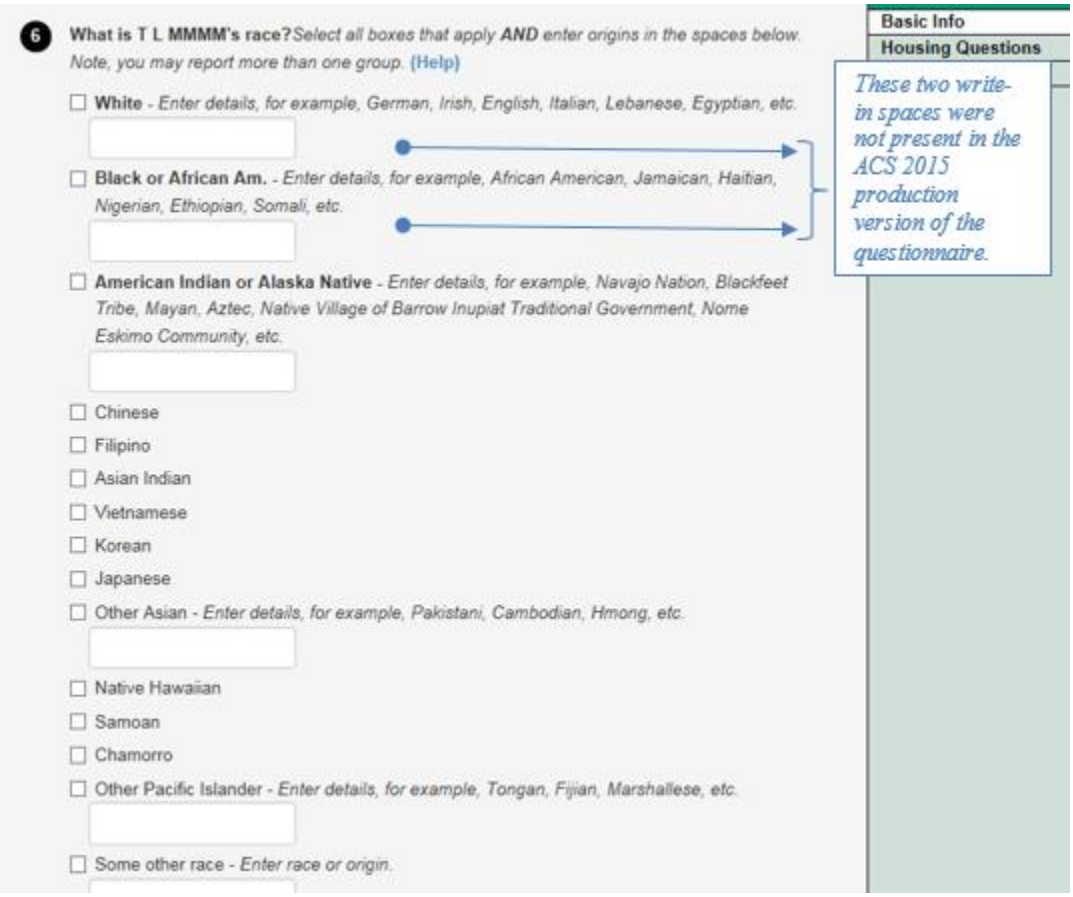


Figure A-3. Image of the Internet Version of the Test Treatment Combined Race and Hispanic Origin Question

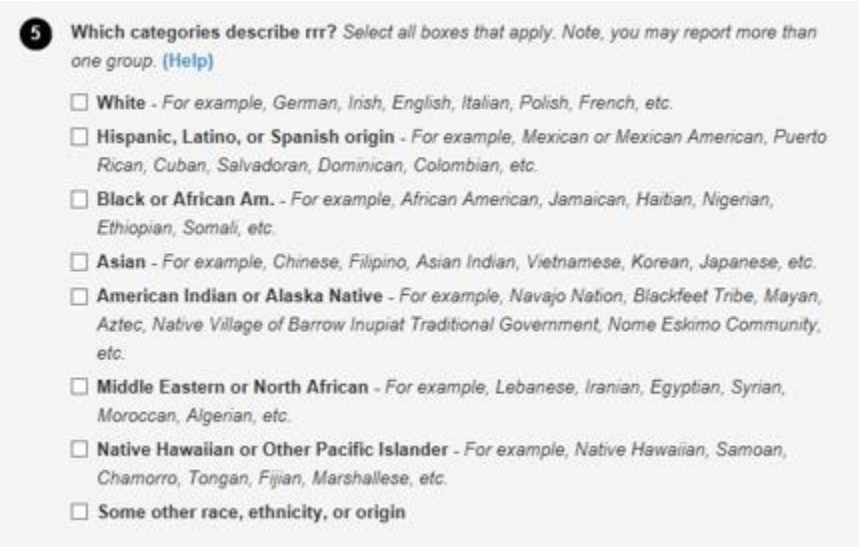


Figure A-4. White Detailed Question for the Test Treatment Race and Hispanic Origin Question

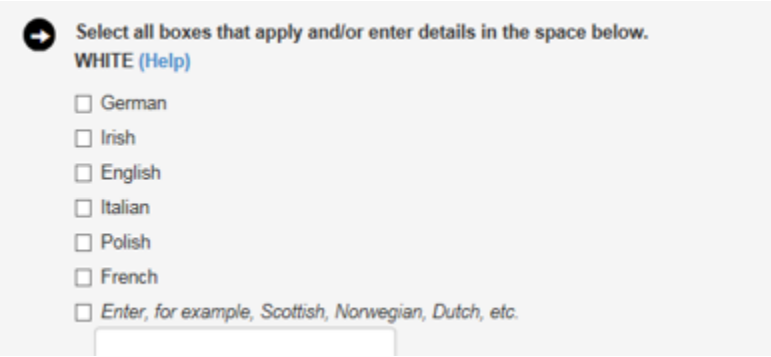


Figure A-5. Hispanic Detailed Question for the Test Treatment Race and Hispanic Origin Question

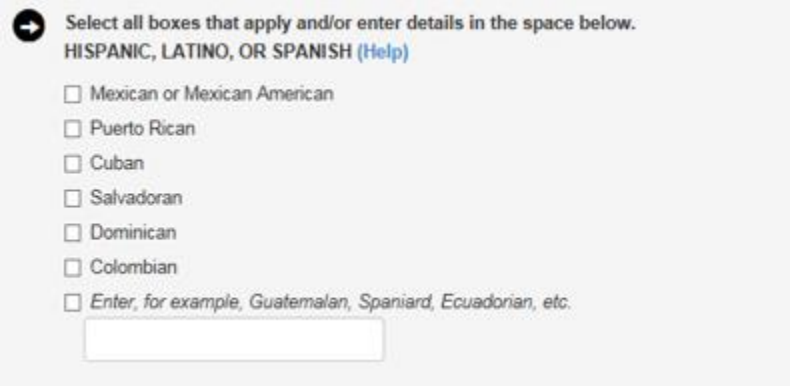


Figure A-6. Black Detailed Question for the Test Treatment Race and Hispanic Origin Question

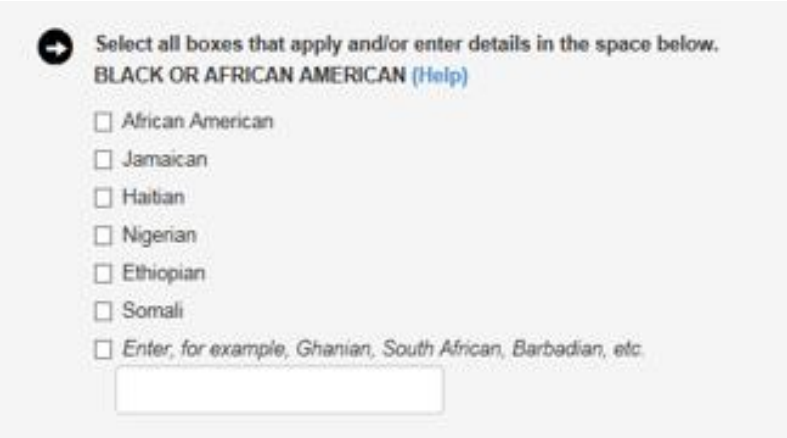


Figure A-7. Asian Detailed Question for the Test Treatment Race and Hispanic Origin Question

Figure A-8. AIAN Detailed Question for the Test Treatment Race and Hispanic Origin Question

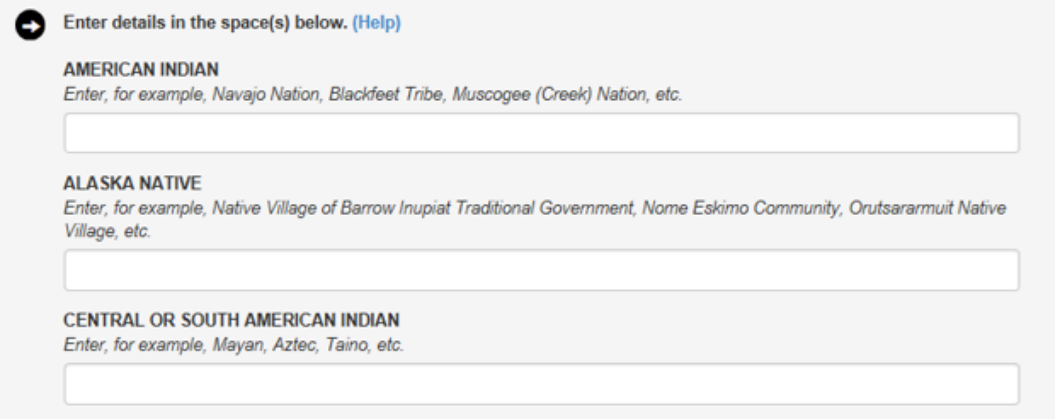


Figure A-9. MENA Detailed Question for the Test Treatment Race and Hispanic Origin Question

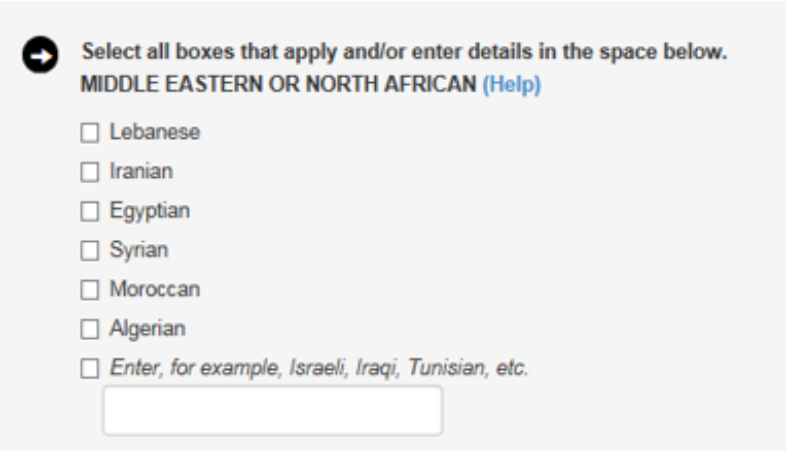


Figure A-10. NHPI Detailed Question for the Test Treatment Race and Hispanic Origin Question

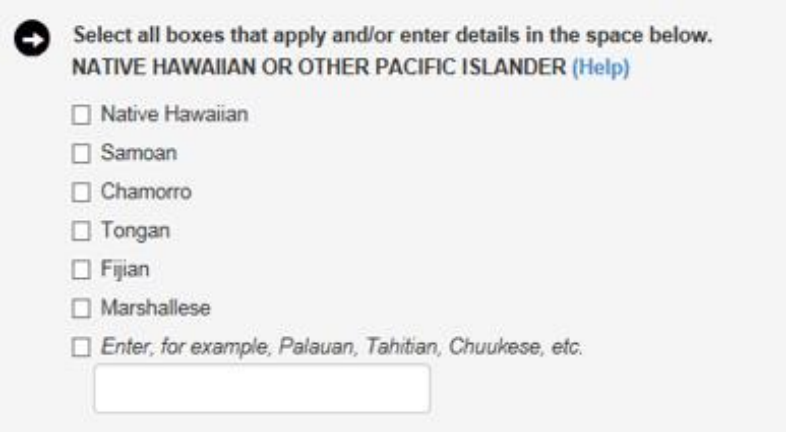
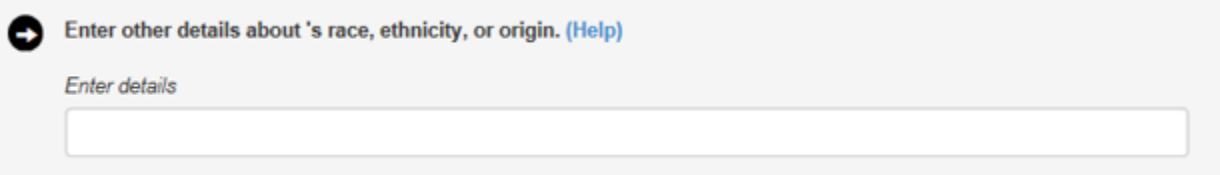


Figure A-11. SOR Detailed Question for the Test Treatment Race and Hispanic Origin Question



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

Control Version	Test Version
[HISPANIC] **Are you of Hispanic, Latino, or Spanish origin?** _Yes – IF YES, ASK HISPANIC DETAIL _No **[HISPANIC DETAIL]** **Are you Mexican, Mexican American, Chicano; Puerto Rican; Cuban, or of some other Hispanic, Latino, or Spanish origin?** _Mexican, Mexican American, or Chicano _Puerto Rican _Cuban _Another Hispanic, Latino, or Spanish origin– IF THIS CATEGORY IS SELECTED, ASK OTHER HISPANIC DETAIL **[OTHER HISPANIC DETAIL]** **What is that origin or origins?** **(For example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.)** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) **[RACE]** **I'm going to read a list of races. You may choose one or more races. For this survey, Hispanic origin is not a race.** **What is your race?** **Are you White; Black or African American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; or Some other race?** _White _Black or African American _American Indian or Alaska Native _Asian _Native Hawaiian or Other Pacific Islander _Some other race	**[RACE/ETHNICITY]** **I'm going to read a list of categories. You may choose all that apply.** **Are you White; Hispanic, Latino, or Spanish origin; Black or African American; Asian; American Indian or Alaska Native; Middle Eastern or North African; Native Hawaiian or Other Pacific Islander; or Some other race, ethnicity, or origin?** _White _Hispanic, Latino, or Spanish Origin _Black or African American _Asian _American Indian or Alaska Native _Middle Eastern or North African _Native Hawaiian or Other Pacific Islander _Some other race, ethnicity, or origin **You said that you are (**NAME THE GROUP OR GROUPS**).** **Now I'm going to collect detailed information. You may give more than one response.**

IF WHITE WAS SELECTED ASK WHITE DETAIL **[WHITE DETAIL]** **What are your WHITE origin or origins? For example, German, Irish, English, Italian, Lebanese, Egyptian, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF BLACK OR AFRICAN AMERICAN WAS SELECTED ASK BLACK DETAIL **[BLACK DETAIL]** **What are your BLACK OR AFRICAN AMERICAN origin or origins? For example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF AMERICAN INDIAN OR ALASKA NATIVE WAS SELECTED ASK AIAN DETAIL **[AIAN DETAIL]** **What are your AMERICAN INDIAN OR ALASKA NATIVE enrolled or principal tribe or tribes? For example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.** _____ (INTERVIEWER TYPES IN TRIBE OR TRIBES) IF ASIAN WAS SELECTED ASK ASIAN DETAIL **[ASIAN DETAIL]** **You may choose one or more Asian groups. Are you Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, or of some other Asian origin?**	IF WHITE WAS SELECTED, ASK WHITE DETAIL **[WHITE DETAIL]** **What are your specific categories for WHITE? For example, German, Irish, English, Italian, Polish, French, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF HISPANIC WAS SELECTED, ASK HISPANIC DETAIL **[HISPANIC DETAIL]** **What are your specific categories for HISPANIC, LATINO, OR SPANISH origin? For example, Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, Colombian, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF BLACK OR AFRICAN AMERICAN WAS SELECTED, ASK BLACK DETAIL **[BLACK DETAIL]** **What are your specific categories for BLACK OR AFRICAN AMERICAN? For example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF ASIAN WAS SELECTED, ASK ASIAN DETAIL **[ASIAN DETAIL]** **What are your specific categories for ASIAN? For example, Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, etc.**

Control Version	Test Version
_Chinese _Filipino _Asian Indian _Vietnamese _Korean _Japanese _Other Asian – IF SELECTED ASK DETAILED ASIAN **[DETAILED ASIAN]** **What is that other Asian origin or origins? (For example, Pakistani, Cambodian, Hmong, etc.)** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF NATIVE HAWAIIAN OR PACIFIC ISLANDER WAS SELECTED ASK NHPI DETAIL: **[NHPI DETAIL]** **You may choose one or more Pacific Islander groups.** **Are you Native Hawaiian, Samoan, Chamorro, or of some other Pacific Islander origin?** _Native Hawaiian _Samoan _Chamorro _Other Pacific Islander – IF SELECTED ASK DETAILED PACIFIC ISLANDER [DETAILED PACIFIC ISLANDER] **What is that other Pacific Islander origin or origins? (For example, Tongan, Fijian, Marshallese, etc.)** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS)	_____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF AMERICAN INDIAN OR ALASKA NATIVE WAS SELECTED, ASK AIAN DETAIL **[AIAN DETAIL]** **What are your specific categories for AMERICAN INDIAN OR ALASKA NATIVE? For example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.** _____ (INTERVIEWER TYPES IN TRIBES, VILLAGES, ETC.) IF MIDDLE EASTERN OR NORTH AFRICAN WAS SELECTED, ASK MENA DETAIL **[MENA DETAIL]** **What are your specific categories for MIDDLE EASTERN OR NORTH AFRICAN? For example, Lebanese, Iranian, Egyptian, Syrian, Moroccan, Algerian, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS) IF NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER WAS SELECTED, ASK NHPI DETAIL **[NHPI DETAIL]** **What are your specific categories for NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER? For example, Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, Marshallese, etc.** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS)

Control Version	Test Version
IF “SOME OTHER RACE OR ORIGIN” WAS SELECTED ASK SOR DETAIL **[SOR DETAIL]** **What is your other race group or groups?** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS)	IF SOME OTHER RACE WAS SELECTED, ASK SOR DETAIL **[SOR DETAIL]** **What are your specific categories for SOME OTHER RACE, ETHNICITY, OR ORIGIN?** _____ (INTERVIEWER TYPES IN ORIGIN OR ORIGINS)

Appendix C: Unit Response Rates Supplemental Table

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

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