Race and Ethnicity Results from the 2016 ACS Content Test

Jacquelyn Harth, Angela Buchanan, Derek Breese, Hyon B. Shin, Merarys Rios, Population Division; Sarah K Heimel, Lindsay Longsine, Decennial Statistical Studies Division; U.S. Census Bureau
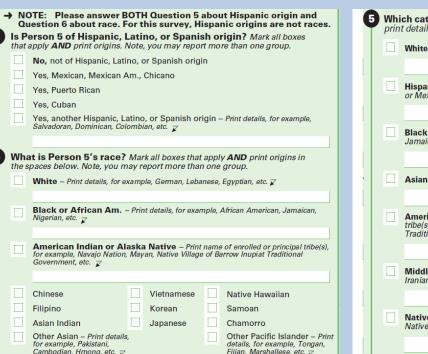
INTRODUCTION

Over the past decade, the Census Bureau has implemented research to address known issues with Hispanic origin and race reporting, as well as concerns raised by data users and community organizations. Prior to the 2010 Census, Census Bureau researchers identified potential issues with the collection of data on Hispanic origin and race. For example, a growing number of respondents chose the response option of Some Other Race (SOR), which was meant to be a residual category. The 2010 Alternative Questionnaire Experiment (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.

The 2016 American Community Survey (ACS) Content Test served as an operational test of the concepts that were investigated in the 2015 National Content Test (NCT). The NCT was the main vehicle for testing newly designed Hispanic origin and race questions. 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.

METHODS

Control Version of Race and Hispanic Origin Questions (Mail)

Test version of Race and Hispanic Origin Question (Mail)

Half of the 2016 ACS Content Test 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.
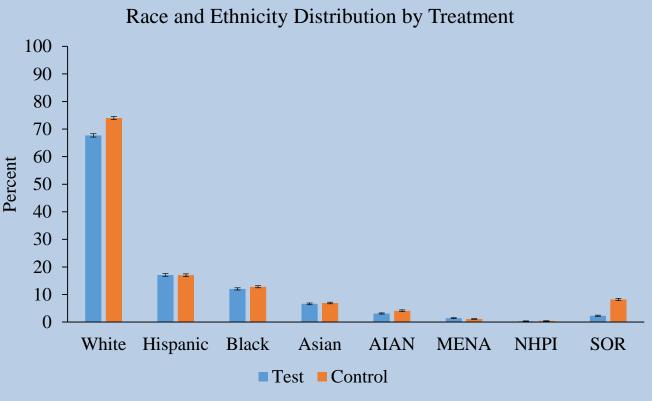
In both treatments, detailed races and ethnicities were collected from all categories through either a write-in line or a checkbox. Within each treatment, the internet, mail, Computer-Assisted Telephone Interview (CATI), and Computer-Assisted Personal Interview (CAPI) modes were all designed similarly, except that the internet-mode design for the test treatment provided six checkboxes to collect detailed race and ethnicities from all major groups in a follow-up screen.
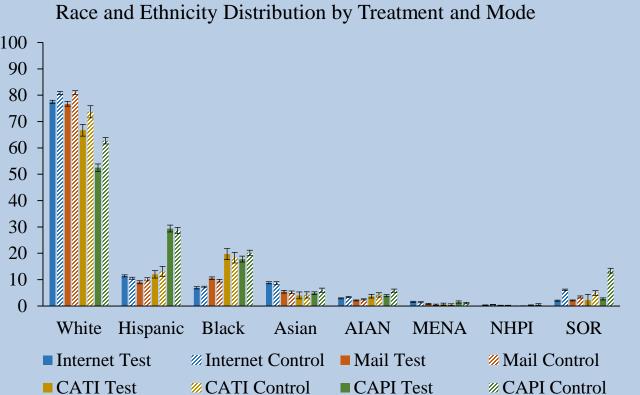
DATA

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 data collection consisted of the following data collection operations: 1) a six week mailout period; 2) a one month CATI period; 3) a one month CAPI period; and 4) a Content Follow-Up telephone reinterview.

The significance level for all hypothesis tests is α = 0.1; all hypothesis tests in this report are two-tailed. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method.

RESPONSE DISTRIBUTIONS

Race and Ethnicity Distribution by Treatment

Race and Ethnicity Distribution by Treatment and Mode

Source: 2016 ACS Content Test

- Overall response distributions were lower in the test treatment for White, American Indian or Alaska Native (AIAN), and SOR respondents.
- There were no differences in reporting for the Hispanic, Black, Asian, MENA, or Native Hawaiian or Other Pacific Islander (NHPI) categories.
- 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 origin.
- 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 trend was observed across modes, with two exceptions:
  - The proportion of White responses was not different between treatments in the CATI mode.
  - The proportion of AIAN responses was only different between treatments in the CAPI mode.
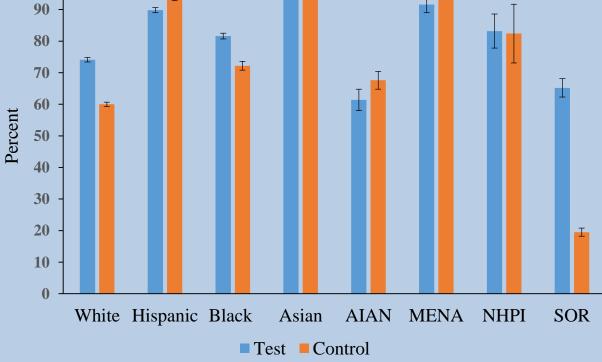
DETAILED REPORTING

- Detailed reporting is defined as 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).
- Overall, total detailed reporting was higher in the test treatment.
- Detailed reporting was higher in the test treatment for respondents identifying as White, Black, or SOR.
- There were no differences in detailed reporting for AIAN, MENA, or NHPI.
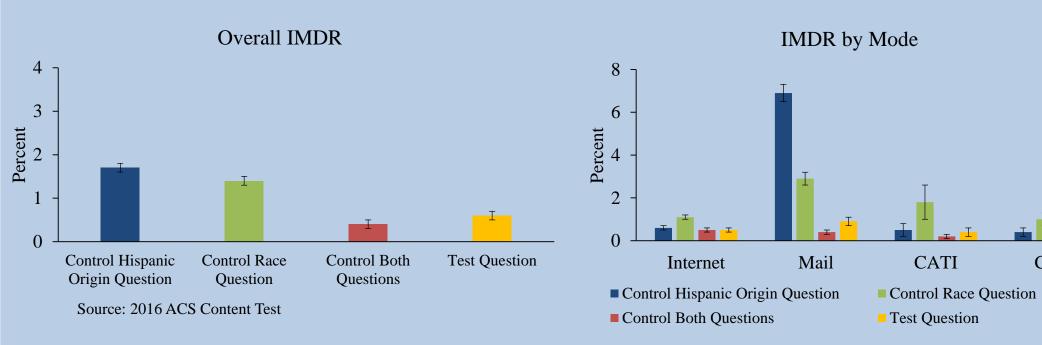- For Hispanic and Asian, the proportion reporting detailed responses was higher in the control.
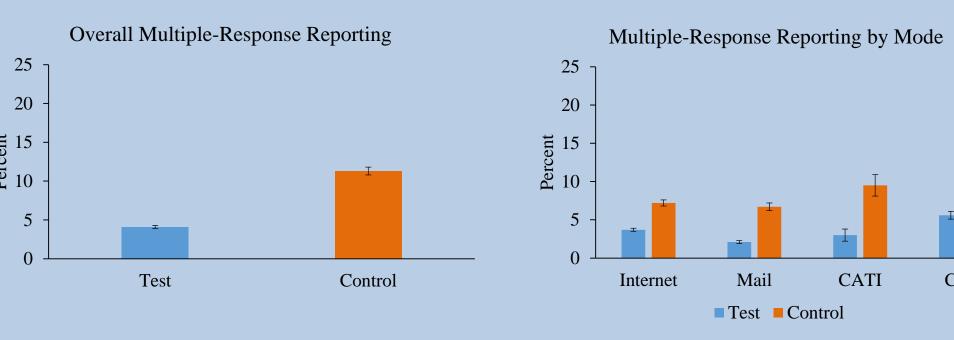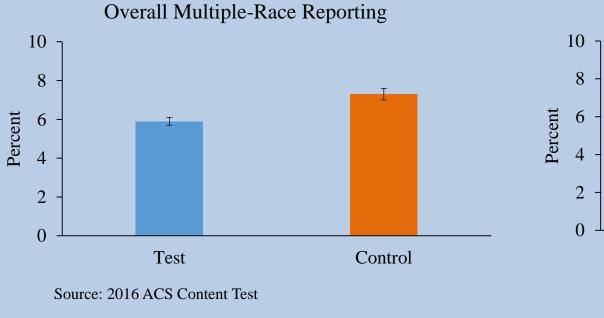

Detailed Reporting

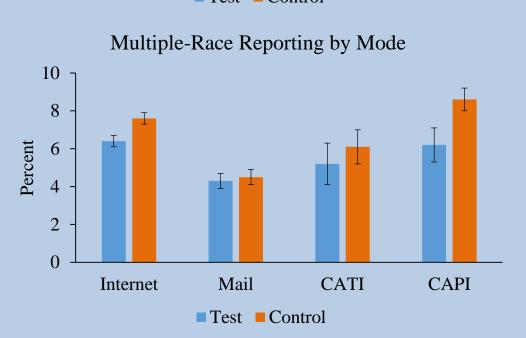Source: 2016 ACS Content Test

ANALYSIS METRICS

- Item missing data rates (IMDR) were lower for the test treatment when compared with the race question in the control treatment.
- When comparing IMDRs in the test treatment to both the race and Hispanic origin questions in the control treatment, no differences were detected.

Overall IMDR

Source: 2016 ACS Content Test

IMDR by Mode

- Multiple-response reporting was lower in the test treatment overall and across all modes.
- 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.
- Similarly, multiple-race reporting was also higher in the control treatment compared with the test treatment overall and for the internet and CAPI modes.
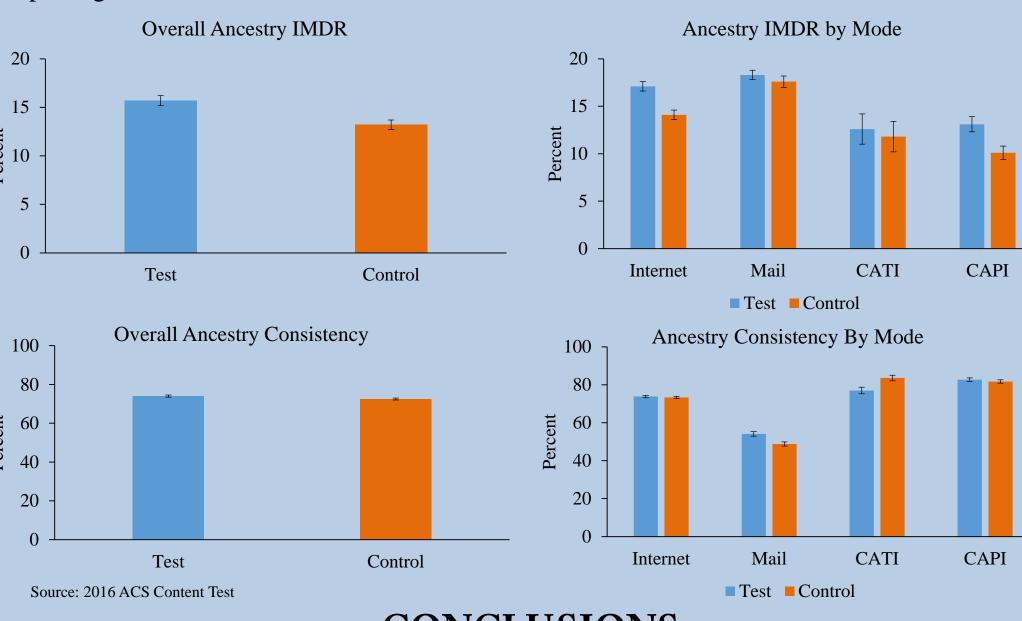

Overall Multiple-Response Reporting

Multiple-Response Reporting by Mode

Overall Multiple-Race Reporting

Multiple-Race Reporting by Mode

Source: 2016 ACS Content Test

DEMOGRAPHIC RESULTS

- White and SOR reporting were lower in the test treatment for every age, sex, education, and housing tenure group.
- AIAN reporting was lower in the test treatment for those under age 18.
- NHPI reporting was lower in the test treatment for those with a high school degree or less education.
- 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.

ANCESTRY RESULTS

- Overall, the ancestry IMDR was higher for the test treatment compared with the control treatment.
- IMDRs for ancestry in the internet and CAPI modes were also higher in the test treatment.
- Overall, respondents in the test treatment reported their race and Hispanic origin more consistently with ancestry than those in the control treatment.
- 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.

Overall Ancestry IMDR

Ancestry IMDR by Mode

Overall Ancestry Consistency

Ancestry Consistency By Mode

Source: 2016 ACS Content Test

CONCLUSIONS

Prior research has shown that when answering a combined question, Hispanic respondents 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 with a MENA category.

Report: Harth, J., Buchanan, A., Breese, D., Shin, H.B., Rios, M., Heimel, S.K., and Longsine, L. (2017). 2016 American Community Survey Content Test Evaluation Report: Race and Hispanic Origin. U.S. Census Bureau, Washington, D.C.