2015 NATIONAL CONTENT TEST
STUDY PLAN

RACE & ETHNICITY

PRE-DECISIONAL:

DO NOT DISTRIBUTE

Nicholas Jones
Population Division

Michael Bentley
Decennial Statistical Studies Division

Sarah Konya
Decennial Statistical Studies Division

Beverly Pratt
Population Division

Version 4

Date 2/10/2016

[This page intentionally left blank]

TABLE OF CONTENTS

1. INTRODUCTION 7

2. BACKGROUND 9
- **2.1 Major Census Tests Over the Past 40 Years** 12
- **2.2 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment** 13
- **2.3 Current Data on Race and Ethnicity** 14

3. METHODOLOGY 15
- **3.1 Research Dimensions for Race and Ethnicity** 16
- **3.2 Overview of the 2015 National Content Test (NCT)** 28
- **3.3 Sample Design** 29
- **3.4 Reinterview Evaluation** 31
- **3.5 Data Processing** 32
- **3.6 Variance Estimation** 35
- **3.7 Introduction to Data Analysis and Decision Criteria for Content Decisions** 35
- **3.8 Testing Alternative Question Formats (Separate vs. Combined)** 36
- **3.9 Testing a Middle Eastern or North African Category** 54
- **3.10 Testing Alternative Instructions and Terminology** 69
- **3.11 Testing Performance of Questions in Paper and Web-Based Designs** 80

4. LIMITATIONS 94

5. MILESTONE SCHEDULE 94

6. DOCUMENT LOGS AND VERSION HISTORY 96

7. REFERENCES 96

APPENDIX A | 2015 NCT WEB-BASED QUESTION DESIGNS 98

APPENDIX B | 2015 NCT PAPER-BASED QUESTION DESIGNS 190

APPENDIX C | 2015 NCT REINTERVIEW QUESTIONS 198

APPENDIX D | 2015 NCT RACE, ETHNICITY, OR ORIGIN HELP TEXT 206

APPENDIX E | CENSUS TRACT-LEVEL ALLOCATION OF 2015 NCT SAMPLE FOR SIX RACE/ETHNIC GROUP STRATA 209

LIST OF TABLES

Table 1. Contact Strategy Panel Design. 28

Table 2. Estimates of Race/Ethnicity Strata in the 2015 NCT Sample. 30

Table 3. 2015 NCT Race/Ethnicity Sample Allocation. 30

Table 4. Percentage of Consistent Responses between Self-Response and Reinterview for Major Race/Ethnic Groups by Question Format. 39

Table 5. Reinterview Race Reporting Compared to Self-Response for Each Question Format. 40

Table 6. Race/Ethnic Group Distribution by Treatment. 42

Table 7. Distribution of Hispanic Responses and Non-Hispanic Responses by Treatment. 42

Table 8. Reporting of Multiple-Responses in Different Question Formats. 43

Table 9. Overall Consistency of Self-Response and Reinterview for Multiple-Responses by Question Format. 44

Table 10. Percentage of Consistent Responses between Self-Response and Reinterview for Selected Multiple-Response Groups by Question Format. 44

Table 11. Self-Response Compared to Reinterview Response for Selected Multiple-Response Groups.45

Table 12. Reporting Patterns of the Hispanic Reinterview Population to the Survey Question Formats.47

Table 13. Pattern of Response by Question Format. 47

Table 14. Detailed Reporting for Major Race/Ethnic Groups by Treatment. 48

Table 15. Reporting of Detailed Groups by Treatment. 49

Table 16. Reporting of "Larger" Detailed Groups and "Smaller" Detailed Groups by Treatment. 52

Table 17. Race Distribution for Hispanic Respondents by Treatment. 53

Table 18. Reporting Patterns of the MENA Reinterview Population to the Survey Question Formats. 58

Table 19. Percentage of Consistent Responses between Self-Response and Reinterview for Major Race/Ethnic Groups by Inclusion or Exclusion of Distinct MENA Category. 59

Table 20. Reinterview Race Distribution Compared to Self-Response. 60

Table 21. Reporting of MENA Responses by Treatment. 61

Table 22. Percentage of Consistent Multiple Responses between Self-Response and Reinterview for MENA Groups by Inclusion or Exclusion of Distinct MENA Category. 62

Table 23. Reporting of Detailed MENA Groups in Different Category Response Areas for Treatments with Distinct MENA Category. 64

Table 24. Reporting of Detailed MENA Groups in Different Category Response Areas for Treatments with No MENA Category. 64

Table 25. Reporting of Detailed MENA Groups by Treatment. 66

Table 26. Race/Ethnic Group Distribution by Treatment. 67

Table 27. Pattern of Response by Presence of Distinct MENA Category. 68

Table 28. Overall Consistency of Self-Response and Reinterview for Multiple-Responses by Question Format. 72

Table 29. Consistency of Self-Response and Reinterview for Major Multiple-Responses Groups by Question Format. 72

Table 30. Reinterview Race Reporting Compared to Self-Response for Major Multiple Response Groups. 73

Table 31. Overall Consistency of Self-Response and Reinterview for Multiple-Responses by Question Format. 74

Table 32. Consistency of Self-Response and Reinterview for Major Multiple-Responses Groups by Question Format. 74

Table 33. Self-Response Race Distribution Compared to Reinterview Race Distribution. 75

Table 34. Consistency of Self-Response and Reinterview for Major Response Groups by OLD Instructions vs. NEW Instructions Approach. 77

Table 35. Consistency of Self-Response and Reinterview for Major Response Groups by Question Terms. 78

Table 36. Detailed Reporting for Major Race/Ethnic Groups by Terms and Instructions. 79

LIST OF FIGURES

Figure 1. OMB Categories and Definitions for Data on Race and Ethnicity 9

Figure 2. Major Census Content Tests Over Past 40 Years..... 12

Figure 3. 2015 NCT Key Dimensions and Research Treatment Paths for Design Testing 17

Figure 4. Summary for 2015 NCT Race, Ethnicity, and Origin Panels 18

Figure 5. Excerpt from 1997 OMB Standards for Race and Ethnicity Guidance on Tabulation 22

Figure 6. Initial Screen for Web-Based Designs..... 25

Figure 7. Subsequent Screen for Web-Based Designs (Hispanic, Latino, or Spanish)..... 26

Figure 8. Subsequent Screen for Web-Based Designs (Asian)..... 27

Figure 9. Range Values for Race, Ethnicity, and Origin Code List 33

Figure 10. Research Questions for Making a Decision Recommendation on Question Format 37

Figure 11. "Larger" Detailed Groups with Dedicated Checkboxes 52

Figure 12. Research Questions for Making a Decision Recommendation on MENA Category 55

Figure 13. Research Questions for Making a Decision Recommendation on Instructions and Terms ... 70

1. INTRODUCTION

The following study plan delineates the execution strategy for race and ethnicity research in the 2015 National Content Test (NCT). First, we lay out the purpose of the 2015 NCT, including a brief overview of the content test. Second, we review relevant literature, with particular emphasis on the 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (AQE). Third, we provide a detailed description on the methodology of the 2015 NCT as it pertains to race and ethnicity. Fourth, we present a detailed description of the research dimensions that will be explored, and a series of research questions, table shells, and decision criteria for making recommendations on the results. Finally, we provide a list of potential limitations, a milestone schedule, document logs, and detailed appendices.

The 2015 NCT will provide the U.S. Census Bureau with information about the content for the race and ethnicity question(s). Results will inform internal planning decisions and will guide the design for the 2020 Decennial Census. We are conducting this mid-decade research in order to make the best decisions possible for the race and ethnicity question(s) on the 2020 Census. Coinciding with this extensive research, we continue with ongoing engagement and discussions about race and ethnicity with the U.S. Office of Management and Budget (OMB), federal statistical agencies, and myriad stakeholder groups. This research will also inform recommendations to the OMB Federal Interagency Working Group (IWG) for Research on Race and Ethnicity, regarding the 1997 OMB Standards and Guidance on Race and Ethnicity. Together, this mid-decade research and community engagement will enable the Census Bureau to provide the most accurate, reliable, and relevant race and ethnicity data possible about our changing and diversifying nation.

The overall objective for the Census Bureau's 2015 NCT race and ethnic research is to test alternative versions of the race and ethnicity questions. This is to gain information about and to improve upon the 2010 Census Hispanic origin and race questions' design and data quality. Our goal is to implement research that refines our efforts to address known race and Hispanic origin reporting issues and important racial and ethnic community concerns while improving data in three crucial areas. These three areas include:

1. Increasing accuracy and reliability of reporting in the major OMB race and ethnic categories,
2. Collecting detailed data for myriad groups, and
3. Obtaining lower item nonresponse rates.

To accomplish this, the 2015 NCT research will evaluate and compare different question designs for race and ethnicity. This will be our primary mid-decade opportunity to compare different decennial content questions prior to making final decisions about the content for the 2020 Census. The 2015 NCT also presents the critical opportunity to compare the success of different question designs to determine how they perform in new web-based data collection methods using the Internet, smartphone, and telephone response options.

Another objective of the 2015 NCT is to test different contact strategies for optimizing self-response. This includes nine different approaches to encourage households to respond and, specifically, to respond using the less costly and more efficient Internet response option. These approaches include altering the timing of the first reminder, use of email as a reminder, altering the timing for sending the mail questionnaire, use of a third reminder, and sending a letter in place of a paper questionnaire to non-respondents. The Census Bureau is committed to using the Internet as a primary response option in the 2020 Census.

The 2015 NCT is part of the research and development cycle leading up to a re-engineered 2020 Census. The test is designed to compare different questionnaire design strategies for key census content areas including race and ethnicity, relationship, and within-household coverage and to provide research for informing recommendations for content decisions. By early 2017, the 2020 Census topics must be submitted to Congress, with the final question wording due in April 2018. That said, the 2015 NCT is our primary mid-decade opportunity to compare different content prior to making final decisions for the 2020 Census. This research will help ensure that the 2020 Census provides the highest-quality statistics about our nation's increasingly changing population.

The 2015 NCT will take place in the late summer of 2015, with a Census Day of September 1. The test will be conducted with a nationally representative sample of 1.2 million housing units in the United States, including Puerto Rico. This sample is designed to ensure that the estimates from this test accurately reflect the nation as a whole, across a variety of demographic characteristics. Related to race and ethnicity, the complex sample design includes oversampling of various race and ethnic groups, including Asian and Native Hawaiian or Other Pacific Islander populations, American Indian or Alaska Natives (AIAN), Black or African Americans, the Hispanic or Latino population, and the Middle Eastern or North African (MENA) population. The NCT is a self-response test only and will not have a nonresponse follow-up component.

Additionally, the 2015 NCT includes a reinterview operation to further assess the accuracy and reliability of the question alternatives for race and ethnicity. The reinterview sample for race and ethnicity includes approximately 75,000 cases. This will enable the Census Bureau to evaluate the key research questions, results, and findings to inform recommendations for the 2020 Census. Through this test, the Census Bureau will also continue testing contact strategies for optimizing self-response, particularly Internet response, building on tests from 2012, 2014, and 2015.

2. BACKGROUND

Since the 1980 Census, the Census Bureau has adhered to federal standards for classifying data collections on race and ethnicity as delineated by the U.S. Office of Management and Budget (OMB). More specifically, since Census 2000, we adhere to OMB's October 30, 1997 "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" for classifying racial and ethnic responses. These standards define five broad categories for data on race and two broad categories for data on ethnicity (for details, see www.whitehouse.gov/omb/fedreg_1997standards).

Figure 1. OMB Categories and Definitions for Data on Race and Ethnicity

OMB CATEGORIES AND DEFINITIONS FOR DATA ON RACE	
American Indian or Alaska Native	"A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment."
Asian	"A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam."
Black or African American	"A person having origins in any of the black racial groups of Africa. Terms such as 'Haitian' or 'Negro' can be used in addition to 'Black or African American.'"
Native Hawaiian or Other Pacific Islander	"A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands."
White	"A person having origins in any of the original peoples of Europe, the Middle East, or North Africa."
OMB CATEGORIES AND DEFINITIONS FOR DATA ON ETHNICITY	
Hispanic or Latino	"A person of Cuban, Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race. The term, 'Spanish origin,' can be used in addition to 'Hispanic or Latino.'"
Not Hispanic or Latino	

OMB standards advise that respondents be offered the option of selecting one or more racial designations. OMB standards also advise that race and ethnicity are two distinct concepts; therefore, Hispanics or Latinos may be of any race. Additionally, OMB standards permit the collection of more detailed information on population groups, provided that any additional groups can be aggregated into the standard broad set of categories.

Data on race and ethnicity have been collected and tabulated in various ways since the first U.S. decennial census in 1790. In fact, Census researchers Karen Humes and Howard Hogan illuminated the complex realities of these changes in their 2009 article, "Measurement of Race and Ethnicity in a Changing, Multicultural America." Humes and Hogan's research provides a historical overview of race and ethnic measurement in U.S. decennial censuses and also provides insights to the ways in which race and ethnicity have been collected and measured over time.

While many respondents report within the race and ethnicity categories specified by OMB standards, it is clear from recent censuses, surveys, and experimental tests that the implementation of the standards is not well understood or is considered unacceptable by a growing number of respondents (Compton et al. 2012; Dowling 2014). This results in respondents' inability or unwillingness to self-identify as OMB standards intended. For a segment of respondents, this arises because of the conceptual complexity that is rooted in OMB standards' definitional distinction between "race" and "Hispanic origin" and the presentation format of the race and Hispanic origin categories.

One key issue is that nearly half of Hispanic or Latino respondents do not identify within any OMB race categories (Rios et al. 2014). With the projected steady growth of the Hispanic or Latino population, the number of people who do not identify with any OMB race categories is expected to increase (Compton et al. 2012; Rios et al. 2014). Another issue is that while the reporting of multiple races is permitted, reporting multiple Hispanic origins or a mixed Hispanic/non-Hispanic heritage in the current Hispanic origin question is not permitted. This differential treatment recognizes interracial unions and multiracial individuals but does not recognize the existence of Hispanic/non-Hispanic unions and individuals or those with a diverse Hispanic heritage.

Coupled with these issues is the reality of what must be done to "modify" reported race data between the decennial census and the development of intercensal population estimates, which serve as the foundation data on race and ethnicity for other federal surveys. The race categories from each Census are reconciled with those race categories that appear in the data from administrative records, which are used to produce population estimates and projections (for details on the modification procedures used in this process for 2010 Census and Census 2000, see www.census.gov/popest/research/modified.html).

This "modification" of the race data did not start in 2010. It has actually been happening for the past several decades. But the issue, and one of the main reasons we undertook the 2010 AQE research to explore alternative measures of race and ethnicity, was in great part because of the recognition that the "modification" of race data was increasing exponentially and becoming a daunting problem that cannot be ignored as it creates a wider and wider disconnect between the full enumeration of the United States population and baseline foundation for other demographic surveys. In the 2010 Census, 19.1 million people (6 percent of all respondents) were classified as Some Other Race alone, and Hispanics made up 97 percent of all those classified as only Some Other Race. Between 2000 and 2010, the population classified as Some Other Race alone increased considerably, growing by about one-quarter in size. In fact, the Some Other Race population has continued to grow since 1980, and was the third largest "race" group overall in 2010, behind the White population and the Black population.

Noting that the Some Other Race category is not an official OMB category, and that it is intended to be a small residual category for respondents who do not identify with any of the minimum OMB race categories, one of the main goals of the AQE was to test designs that would increase reporting within the OMB categories, and reduce the reporting of Some Other Race. One of the most notable AQE findings was that while the separate questions still had Some Other Race as high as 7 percent,

the combined question designs yielded a substantially reduced Some Other Race population under one-half percent. We know from the AQE research that this is largely due to Hispanics choosing their identity (i.e., only "Hispanic") in the combined question format. Overall, when a Hispanic category is provided as a response option, Some Other Race becomes one of the smallest response categories, demonstrating that a combined question approach is more in-line with how Hispanic respondents view themselves. The AQE reinterview study and AQE focus group research confirmed that these reporting patterns were a closer reflection of how Hispanics self-identify, and this was a major finding of the research.

Additionally, prior to the 2010 Census, different racial and ethnic communities lobbied the Census Bureau and the U.S. Congress for additional changes to categories in the race and Hispanic origin questions. This amplified the concerns a number of racial and ethnic communities, such as Middle Eastern populations or Afro-Caribbean populations, have about self-identifying in the OMB standard categories used in decennial census questions and on other federal surveys. The growing lack of understanding or acceptance of the current OMB standards is compounded by:

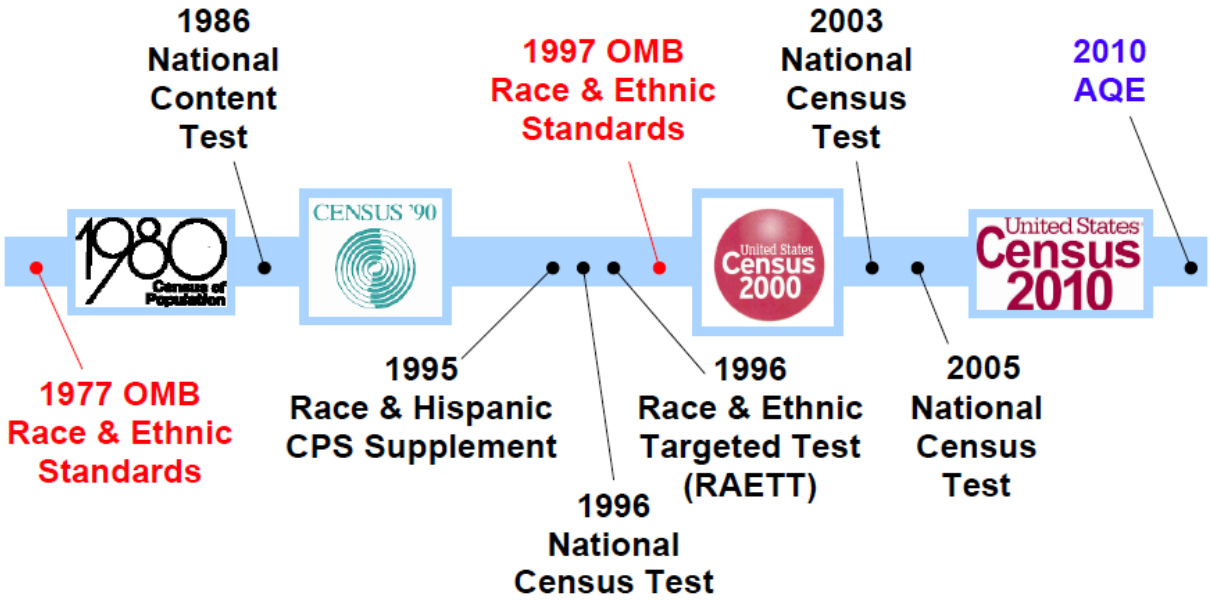
- The rapidly changing demographics of the U.S. population;
- The increase and complexity of immigration flows from all corners of the globe (Newby and Dowling 2007; Roth 2012);
- A fluidity of racial and ethnic self-identification;
- Increasing responses of "Some Other Race" on census surveys; and
- Widespread campaigns and lobbying of the Census Bureau, the OMB, and the United States Congress for changes to the race and Hispanic origin questions and categories.

All of these issues point to the importance of conducting thorough research related to the design of the race and Hispanic origin questions as the Census Bureau embarks upon preparations for the 2020 Census. The Census Bureau will continue working with OMB, other federal statistical agencies, and key external advisors and stakeholders as research is planned and conducted throughout the decade related to alternative approaches to the implementation of the OMB standards in decennial censuses and surveys. From our review of recent social scientific literature, we note there are not many empirical studies, outside of those conducted by the Census Bureau, which analyze formatting of the race and ethnicity question(s), the inclusion of a MENA category, or revisions to examples and question terminology for improving data on race and ethnicity. As such, the present study should help to inform the literature on this important topic.

2.1 | Major Census Content Tests Over Past 40 Years

It is important to note that the 2010 AQE research is one of many of decennial census content tests that were focused on improving race and ethnic data since the 1970s. Census content tests are one of the main mechanisms the Census Bureau uses to develop research questions on the census forms, in an effort to improve the data from decade to decade. Figure 2 below illustrates a history of the major race and ethnic content tests over the past 40 years, with the 2010 AQE being the most recent.

Figure 2. Major Census Content Tests Over Past 40 Years



Note: OMB = U.S. Office of Management and Budget; CPS = Current Population Survey

The Census Bureau remains committed to improving the accuracy and reliability of census results by researching approaches that more accurately measure and reflect how people self-identify their race and ethnicity. This commitment is reflected in numerous past Census Bureau studies, as illustrated above, that have been conducted on race and Hispanic origin reporting (U.S. Census Bureau 1997; Sheppard et. al. 2004; Alberti 2006; Fernández et al. 2009; Childs et al. 2010). Interestingly, both the 1996 RAETT (U.S. Census Bureau 1997) and the 2005 National Content Test (Alberti 2006) demonstrated over the past couple of decades that when presented with separate race and Hispanic origin questions, Hispanics have great difficulty responding to the race question.

As shown in the above figure, the 2015 NCT research builds on extensive research on race and ethnicity previously conducted by the Census Bureau to examine how people in our society identify their race and ethnicity as our society grows more diverse and complex. This research acknowledges that a growing number of people find the current race and ethnic categories confusing, or they wish to see their own specific group reflected on the census. Following this research, the 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (AQE) was fielded as the most comprehensive research effort on race and Hispanic origin ever undertaken by the Census Bureau.

2.2 | 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment

The 2010 AQE research focused on improving the race and Hispanic origin questions by testing a number of different questionnaire design strategies. The primary research objectives of the AQE were to design and test questionnaire strategies to increase reporting in the major OMB race and ethnic categories, elicit reporting of detailed race and ethnic groups, lower item nonresponse rates, and increase accuracy and reliability of the results (Compton et al. 2012). The 2010 AQE was comprised of three components:

1. A mail out/mail back sample with half a million households;
2. A telephone reinterview with one-in-five of those households; and
3. A series of 67 focus groups with about 800 people across the country, including Alaska, Hawaii, and Puerto Rico.

The Census Bureau conducted the 2010 AQE research to better understand how and why people identify themselves in different ways and in different contexts. The 2010 AQE examined alternative question design strategies for improving the collection of data on race and Hispanic origin, with four goals in mind:

1. Eliciting detailed responses for all racial and ethnic communities (e.g., Chinese, Mexican, Jamaican, Lebanese, etc.);
2. Increasing the accuracy and reliability of the results;
3. Increasing responses to the race and ethnicity question(s); and
4. Increasing reporting in the standard race and ethnic categories, as defined by the OMB.

The results of the AQE supported all of these objectives. One of our experimental approaches asked about race and Hispanic origin in one combined question. In the combined question, each major race and ethnic group had a checkbox with examples and a write-in line where respondents could provide detailed responses. Many individuals across communities liked the combined question approach, and felt it presented equity to the different categories. The AQE's results led to some promising strategies to address the challenges and complexities of race and Hispanic origin measurement and reporting. Some of the findings from this research included:

- Combining race and ethnicity into one question did not change the percentage of people who reported as Hispanics, Blacks, Asians, American Indians and Alaska Natives, or Native Hawaiians and Other Pacific Islanders (Hill and Bentley 2014).

- The combined question yielded higher item response rates, compared with separate question approaches.

- The combined question increased reporting of detailed responses for most groups, but decreased reporting for others.

- The combined question more accurately reflected self-identity.

One of the main goals of the AQE was to test designs which would increase reporting in the standard OMB categories, and which therefore would reduce the reporting of Some Other Race. The Some Other Race category is not an official OMB category, as it is intended to be a small residual category for

respondents who do not identify with any of the minimum OMB race categories. Nonetheless, Some Other Race has continued to grow since 1980, into the third largest race group overall. One of the most notable AQE findings was that while the separate questions still had Some Other Race as high as 7 percent, the combined question designs yielded a substantially reduced Some Other Race population under one-half percent. We know from AQE that this is largely due to Hispanics choosing their identity (i.e., only "Hispanic") in the combined question format. Overall, when a Hispanic category is provided as a response option, Some Other Race becomes one of the smallest response categories, demonstrating that a combined question approach is more in-line with how Hispanic respondents view themselves. The 2010 AQE reinterview study and focus group research confirmed that these reporting patterns were a closer reflection of how Hispanics self-identify, and this was a major finding of the research.

The 2010 AQE research marked the beginning of race and ethnicity research for this decade. The research yielded critical findings, from which additional experimental question refinements and new research topics emerged. These research topics evolved over the past several years with ongoing qualitative and quantitative research, through internal discussions among Census Bureau experts, as well as external dialogues with advisors, race and ethnic scholars, OMB, federal statistical and policy agencies, and myriad community leaders and stakeholders. The successful strategies from the AQE research, as well as additional tests in 2012, 2014, and early 2015, have been employed in the design of the Census Bureau's mid-decade research for the 2020 Census. The scope of the 2015 NCT builds upon the successful strategies of the Census Bureau's 2010 AQE research, and examines several dimensions for improving data on race and ethnicity, each of which will be discussed in detail later in this study plan:

- Question format, including evaluating performance of paper-based questions and new web-based data collection methods
- Response categories
- Wording of instructions
- Question terminology

2.3 | Current Data on Race and Ethnicity

Over the last few decades, many Census Bureau studies have examined race reporting among Hispanics on the census questionnaire, but these studies did not specifically look at those who self-reported being of Hispanic origin. In March 2014, the Census Bureau released a blog by Population Division researchers Merarys Ríos-Vargas and Fabián Romero, titled, "*Shedding Light on Race Reporting Among Hispanics*" (see http://blogs.census.gov/2014/03/28/shedding-light-on-race-reporting-among-hispanics/).

The authors' research, "*Race Reporting Among Hispanics: 2010*," examined this topic and found that more than two-fifths (43.5 percent) of self-reported Hispanics did not report belonging to any federally recognized race group as defined by OMB. This includes 30.5 percent who reported or were classified as "Some Other Race" (SOR) only. Respondents are classified this way when they only check and/or write-in responses not categorized as any of the OMB race groups. An additional 13.0 percent of self-reported Hispanics did not provide a response to the race question. The findings from this study are intended to supplement the results presented in the 2010 AQE report (for more details, see www.census.gov/population/www/documentation/twps0102/twps0102.pdf).

Currently, the Census Bureau collects additional detailed information on Hispanic or Latino groups, American Indian and Alaska Native tribes, Asian groups, and Native Hawaiian and Other Pacific Islander groups. For example, responses to the race question such as Navajo Nation, Nome Eskimo Community, and Mayan are collected and tabulated separately in Census Bureau censuses and surveys, but also are aggregated and tabulated into the total American Indian or Alaska Native population. Similarly, responses to the race question such as Chinese, Asian Indian, and Vietnamese are collected and tabulated separately, but also aggregated and tabulated into the total Asian population, while responses such as Native Hawaiian, Chamorro, or Fijian are collected and tabulated separately, but also tabulated and aggregated into the total Native Hawaiian or Other Pacific Islander population. Responses to the ethnicity question such as Mexican, Puerto Rican, and Cuban are collected and tabulated separately, but also tabulated and aggregated in Census Bureau censuses and surveys, and into the total Hispanic or Latino population.

The 2015 NCT will test ways to collect and tabulate detailed information for the detailed groups, not just to the broad groups, such as Asian or Hispanic. Detailed data for specific White population groups, such as German, Irish, and Polish, and specific Black population groups, such as African American, Jamaican, and Nigerian, will be collected and tabulated, and can be aggregated, respectively, into the total White population and the total Black population.

The 2015 NCT also includes testing of a separate MENA category and the collection of data on detailed MENA groups, such as Lebanese, Egyptian, and Iranian. Currently, following the 1997 OMB standards, MENA responses are classified under the White racial category, per OMB's definition of "White" (refer to Figure 1 for the specific definition).

We are working to address many of the questions and concerns from myriad racial and ethnic communities through outreach and engagement about our mid-decade testing, discussion of new explorations that have the advantage of web-based technology to collect data on major groups as well as detailed groups, and even the retention (and potential creation) of detailed checkboxes on various question designs. All of our research is working towards the broader goal of balance and equality across communities for the opportunity to self-identify their race/ethnicity and receive the return of critical data for both long-standing groups and recently emerging groups in the United States.

3. METHODOLOGY

Qualitative research and field-testing, such as focus groups and cognitive testing, are critical components for understanding ways to improve quantitative data on race and ethnicity, and they are essential for obtaining information on how well revised experimental race and ethnicity questions perform when asked of respondents. As refinements and tweaks to the experimental race and ethnicity questions have been made over the past several years, qualitative research was conducted to provide insight on whether respondents understand the revised questions correctly and if they provide answers that reflect their "true" self-identification. These efforts aim to remedy aspects of the questions that are misunderstood by respondents or are problematic, prior to fielding the 2015 NCT. Once quantitative data have been collected via the 2015 NCT field test, analyses of both the quantitative and qualitative results can provide critical information on respondent reporting patterns, document potential sources of respondent difficulty, as well as provide a richer understanding of the data that have been collected.

3.1 | Research Dimensions for Race and Ethnicity

As previously mentioned, the 2015 NCT research will examine several key dimensions for improving the data on race and ethnicity. The following sections will describe each of the key dimensions in detail:

- Question format, including evaluating performance of paper-based questions and new web-based data collection methods
- Response categories
- Wording of instructions
- Question terminology

3.1.1 | Description of Race and Ethnicity Treatments

The flowchart on the next page (see Figure 3) shows each of the dimensions being tested in the 2015 NCT, including both web-based designs and paper-based designs. The key dimensions are:

- Separate question vs. combined questions (question format)
- MENA vs. No MENA (response categories),
- Mark [***X***] one or more boxes vs. Mark all that apply (instruction wording)
- Race/Origin vs. Race/Ethnicity vs. using no terms – "categories" (question terminology)

There are 36 different web-based panels, labeled 1 through 36. The flowchart compares the "Combined + write-in areas" and the "Combined + 6 checkboxes & write-ins." The "Combined + write-in areas" is similar to the 2014 Census Test Internet version where a write-in area follows each of the major race/ethnic categories. The "Combined + 6 checkboxes & write-ins" is our new design for the 2015 NCT, where a series of detailed checkboxes and a write-in area are employed. Images of the web-based versions (1 through 36) can be found in Appendix A.

Eight different versions have been developed for paper, labeled A, C, D1, D2, G, H, I, and W. These paper versions are labeled in Figure 3 to show the connections for how paper versions match the web-based versions. Images of these versions can be found in Appendix B). The goal of our research is to test the key dimensions in new web-based designs while also ensuring that these dimensions are researched on traditional paper data collection modes. We are testing the fully factorial design of the web-based panels and have included selected paper treatments that correspond with the main differences across the key dimensions. While originally planning 12 treatments, operational constraints limited us to eight paper panels. Please see Figure 3 and Figure 4 for more information.

Figure 3. 2015 NCT Key Dimensions and Research Treatment Paths for Design Testing

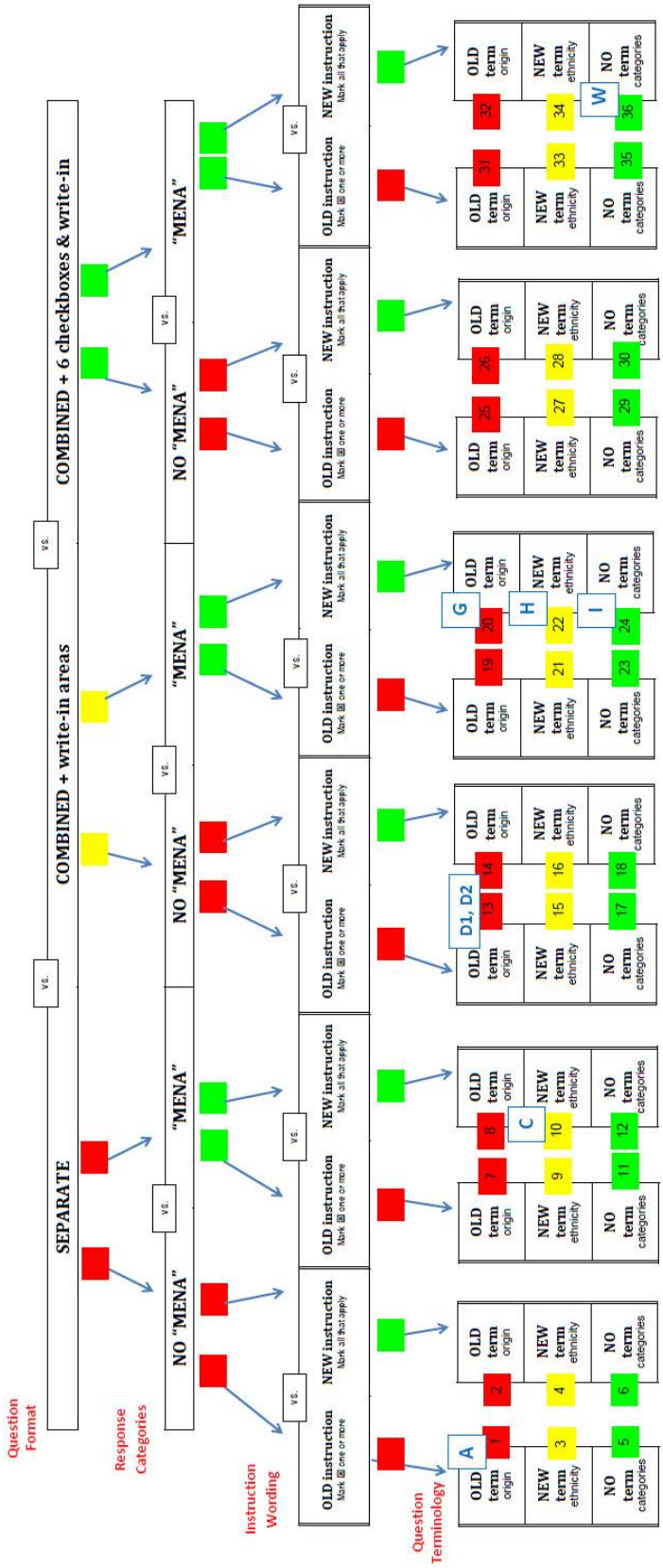


Figure 4. Summary of 2015 NCT Race, Ethnicity, and Origin Panels

VERSION	QUESTION FORMAT			INCLUSION OF "MENA"		INSTRUCTIONS		TERMINOLOGY		
	Separate	Combined + Write-ins	Combined + Checkboxes	No "MENA"	"MENA"	Old	New	Old = "origin"	New = "ethinicity"	No term = "categories"
1/A	X			X		X		X		
2	X			X			X	X		
3	X			X		X			X	
4	X			X			X		X	
5	X			X		X				X
6	X			X			X			X
7	X				X	X		X		
8	X				X		X	X		
9	X				X	X			X	
10/C	X				X		X		X	
11	X				X	X				X
12	X				X		X			X
13/D1, D2		X		X		X		X		
14		X		X			X	X		
15		X		X		X			X	
16		X		X			X		X	
17		X		X		X				X
18		X		X			X			X
19		X			X	X		X		
20/G		X			X		X	X		
21		X			X	X			X	
22/H		X			X		X		X	
23		X			X	X				X
24/I		X			X		X			X
25			X	X		X		X		
26			X	X			X	X		
27			X	X		X			X	
28			X	X			X		X	
29			X	X		X				X
30			X	X			X			X
31			X		X	X		X		
32			X		X		X	X		
33			X		X	X			X	
34			X		X		X		X	
35			X		X	X				X
36/W			X		X		X			X

Analogous with Figure 3, Figure 4 (shown above) demonstrates how the 36 web-based panels (along with the matching lettered paper panels) corresponds to the key dimensions being explored: 1) question format, 2) response categories, 3) instruction wording, and 4) terminology.

3.1.2 | Scope, Objectives, and Question Designs for Testing Race and Ethnicity Content in the 2015 NCT

The scope of the 2015 NCT builds upon the successful strategies from the 2010 AQE research, and undertakes further testing to examine several key dimensions for the questions on race and ethnicity. Each of the key research dimensions are presented in detail below, along with a description of relevant question designs that are being tested in the 2015 NCT.

One dimension is **question format** – as we continue to research the separate questions approach and the combined question approach. This dimension includes the overarching comparison of **paper-based question designs** and **web-based question designs** – with the advantage of technology, such as the Internet, smartphones, and Telephone Questionnaire Assistance (TQA), to enhance question designs and optimize reporting of detailed racial and ethnic groups.

Another dimension examines the **response categories** – by exploring how to collect and tabulate data for respondents of Middle Eastern and North African heritage in the United States.

Additionally, we have a dimension which pertains to the **wording of instructions** , as well as a dimension that focuses on **question terminology** – through examining ways to optimize detailed reporting and to improve respondent understanding of the options to report multiple race and ethnic groups.

Question Format Dimension. The 2015 NCT will evaluate the use of two alternative question approaches for collecting detailed data on race and ethnicity. One approach uses two separate questions – the first about Hispanic origin, and the second about race. The other approach combines the two items into one question about race and ethnicity. The 2015 NCT research will test both approaches with new web-based data collection methods, including Internet, telephone, and in-person response. Each approach is described below, along with its associated data collection mode(s) (i.e., paper-based question designs and/or web-based question designs).

1	**Separate questions for race and for Hispanic origin** **(paper and web-based)**	This is a modified version of the race and Hispanic origin approach used in the 2010 Census. Revisions based on the 2010 AQE research include adding write-in areas and examples for the White response category and for the Black or African American response category, removal of the term "Negro," and the addition of an instruction to allow for multiple responses in the Hispanic origin question. *Note: Refer to Appendix A and B. (Separate Questions)*
2	**Combined question with checkboxes and write-ins visible at same time** **(paper)**	This is a modified version of the combined question approaches found to be successful in the 2010 AQE research. Checkboxes are provided for the major race and ethnic categories, with a corresponding write-in space for detailed responses to each checkbox category. In this version, all checkboxes and write-in spaces are visible at all times. Each response category contains six example groups, which represent the diversity of the geographic definitions of the respective OMB category. For instance, the Asian category employs examples of Chinese, Filipino, Asian Indian, Vietnamese, Korean, and Japanese, which represent the six largest detailed Asian groups in the United States, reflecting OMB's definition of Asian ("A person having origins in any of the original peoples of the Far East, Southeast Asia, and the Indian subcontinent."). Respondents do not have to select a major checkbox, and may enter a detailed response in the write-in space without checking a category. *Note: Refer to Appendix B. (Streamlined)*

3	**Combined question with major checkboxes, detailed checkboxes, and write-ins** **(paper)**	This is a modified version of the combined question approaches found to be successful in the 2010 AQE. Checkboxes are provided for the major race and ethnic categories, along with a series of detailed checkboxes under each major category, and a corresponding write-in space and examples to elicit and collect all other detailed responses within the major category. In this version, all checkboxes and write-in spaces are visible at all times. Again, the detailed response categories represent the diversity of the geographic definitions of the respective OMB category. For instance, under the Asian category (and major checkbox), a series of detailed checkboxes is presented for Chinese, Filipino, Asian Indian, Vietnamese, Korean, and Japanese, which represent the six largest detailed Asian groups in the United States. Then, instructions to enter additional detailed groups (with the examples of "Pakistani, Thai, Hmong, etc.") precede a dedicated write-in area to collect other detailed Asian responses. Again, these detailed groups reflect OMB's definition of Asian ("A person having origins in any of the original peoples of the Far East, Southeast Asia, and the Indian subcontinent."). Respondents do not have to select a major race/ethnic checkbox, and may enter a detailed response in the write-in area without checking a category. *Note: Refer to Appendix B. (Multiple Detailed Checkboxes)*
4	**Combined question with major checkboxes and write-ins on separate screens** **(web-based)**	In this version, the detailed origin groups are solicited on subsequent screens after the major response categories have been selected on the initial screen. On the initial screen, the major checkbox categories are shown along with their six representative example groups. Once the major categories have been selected, one at a time, subsequent screens solicit further detail for each category that was chosen (e.g., Asian), using a write-in space, with examples, to collect the detailed groups (e.g., Korean and Japanese). The intent is to separate mouse click tasks (checkbox categories) and typing tasks (write-ins) in an attempt to elicit responses that are more detailed. This approach was used as one of three race/ethnicity Internet panels in the 2014 Census Test. *Note: Refer to Appendix A. (Subsequent Write-In Screens)*
5	**Combined question branching with detailed checkbox screens** **(web-based)**	This version is an alternative method of soliciting detailed origin groups using separate screens, detailed checkboxes, and write-in spaces. On the first screen, the major checkbox categories are shown along with their six representative example groups. Once the major categories have been selected, one at a time, subsequent screens solicit further detail for each category, this time using a series of additional checkboxes for the six largest detailed groups (e.g., Chinese, Filipino, Asian, Indian, Vietnamese, Korean, and Japanese) with a write-in space also provided to collect additional groups. *Note: Refer to Appendix A. (Subsequent Detailed Checkbox Screens)*

One benefit of the web-based response mode is that it allows for more functionality and greater flexibility in designing questions compared to paper, which is constrained by space availability. The 2015 NCT will utilize innovative web-based technology, such as the Internet, smartphones, and tablets to improve question designs, and to optimize reporting of detailed racial and ethnic groups (e.g., Samoan, Iranian, Blackfeet Tribe, Filipino, Jamaican, Puerto Rican, Irish, etc.). These web-based designs also provide much more utility and flexibility for using detailed checkboxes and write-in spaces to elicit and collect data for detailed groups than traditional paper questionnaires, and will help collect data for both the broader race/ethnic categories, as well as more detailed responses across all groups.

Comparing the Separate Questions Approach with the Combined Question Approach

There is complexity involved with making comparisons between the results from the separate questions approach and the results from the combined question approach. One of the important measures we will examine is the overall level of item nonresponse for the separate questions approach and for the combined question approach. There is complexity involved with making these comparisons with respect to "item" nonresponse for the different formats, and the critical conceptual question to consider here is, what is considered "item nonresponse."

The separate question approach may have two item nonresponse rates, one for the Hispanic Origin question and one for the Race question; and both will be examined. The combined question format will have one item nonresponse rate for the race/ethnic question.

For the separate question on Hispanic origin, if a respondent does not answer the question at all – no checkboxes are marked, and no write-ins are provided, this would constitute what we conceptualize as "item nonresponse" to the Hispanic origin question. If they answer affirmatively, that they are of Hispanic origin (by selecting one of the Hispanic checkbox categories and/or writing in a term that is classified as "Hispanic") we conceptualize this as a valid response. Similarly, if they report they are "Not of Hispanic origin" (by marking the "No, not of Hispanic origin" checkbox, or writing in a term that is classified as "non-Hispanic") we conceptualize this as a valid response. Additionally, a "Hispanic" and "Non-Hispanic" response is also acceptable, and we conceptualize this as a valid response.

Considering the concept of the combined question approach, where a separate "Hispanic or Latino" category is placed coequally among the other major response categories (White, Black, Asian, AIAN, MENA, NHPI, and SOR), when a respondent does not answer the question at all – no checkboxes are marked, and no write-ins are provided, this constitutes what we conceptualize as "item nonresponse" to the combined question. If they report they are "Hispanic" by marking the "Hispanic" checkbox, or writing in a term that is classified as "Hispanic," we conceptualize this as a valid response.

If they do not report they are "Hispanic" and they report one or more "race" categories (e.g., White, Black, Asian, etc.) then we classify them with the reported race(s) and determine that they are "not of Hispanic origin" as they did not report that they were. This is an important concept to address, as it is line with the approach taken by the OMB in the guidance for a combined question in the 1999 standards. As shown in Figure 5 on the next page (highlighted in green), when using a combined question format, both race (including multiple responses) and ethnicity may be collected through the question, but the selection of only one category (either race or ethnicity) in the combined format is acceptable.

This means that if a respondent only reports "Hispanic" in the combined question, that is acceptable. It also means that if a respondent only reports "Black" in the combined question, that is acceptable. Of course, it is also acceptable for the respondent to report multiple groups, as well as a "race" (e.g., Black) and "ethnicity" (e.g., Hispanic). This premise also extends to the way in which the collected data are presented, as discussed at the bottom of the excerpt from the 1997 standards shown in Figure 5. The "Hispanic" category is presented co-equally among the "race" categories, and all are mutually exclusive. Therefore, a response to the combined question is considered to be complete if the respondent provides at least one of the categories.

Figure 5. Excerpt from 1997 OMB Standards for Race and Ethnicity Guidance on Tabulation

2. Data Formats

b. Combined format

The combined format may be used, if necessary, for observer-collected data on race and ethnicity. Both race (including multiple responses) and ethnicity shall be collected when appropriate and feasible, although the selection of one category in the combined format is acceptable.

If a combined format is used, there are six minimum categories:
-- American Indian or Alaska Native
-- Asian
-- Black or African American
-- Hispanic or Latino
-- Native Hawaiian or Other Pacific Islander
-- White

When aggregate data are presented, data producers shall provide the number of respondents who marked (or selected) only one category, separately for each of the six categories. In addition to these numbers, data producers are strongly encouraged to provide the detailed distributions, including all possible combinations, of multiple responses. In cases where data on multiple responses are collapsed, the total number of respondents reporting "Hispanic or Latino and one or more races" and the total number of respondents reporting "more than one race" (regardless of ethnicity) shall be provided.

This is the conceptual premise we follow as we make comparisons with respect to "item" nonresponse for the different question design formats. This is also a critical conceptual component which must be addressed in the current OMB Standards for Race and Ethnicity, moving forward, if a recommendation is made to employ a combined question approach.

Response Categories Dimension. The 2015 NCT will also evaluate the use of a "Middle Eastern or North African" ("MENA") response category to collect data for respondents of Middle Eastern or North African heritage in the United States. There will be two treatments for testing this dimension:

1	**No separate MENA category**	This treatment tests approaches without a separate MENA checkbox category. Here, the MENA responses and represents the current OMB definition of White. With this approach, the White racial category provides examples of both Middle Eastern and North African origins (e.g., Lebanese; Egyptian) along with examples of European origins (e.g., German; Irish) as part of the currently defined "White" racial category. *Note: Refer to Appendix A, panel #s 1-6, 13-18, and 25-30*
2	**Use of a distinct MENA category**	This treatment tests the addition of a distinct MENA checkbox category for respondents of Middle Eastern or North African heritage in the United States. The MENA category is placed within the current category lineup, based on estimates of population size, between the category for American Indians and Alaska Natives and the category for Native Hawaiians and Other Pacific Islanders. With the addition of this new category, the "White" example groups are revised. The Middle Eastern and North African examples of Lebanese and Egyptian are replaced with the European examples of Polish and French. The MENA checkbox category will have the examples of Lebanese, Iranian, Egyptian, Syrian, Moroccan, and Algerian. All other checkbox categories and write-in spaces remain the same. *Note: Refer to Appendix A, panel #s 7-12, 19-24, and 31-36*

For the purpose of the 2015 NCT, the Census Bureau developed a working classification of the "Middle Eastern or North African" category. The development of this MENA working classification is described in detail in Section 3.9 (Testing a Middle Eastern or North African category), but we introduce it here to provide context for understanding the way in which the category is constructed for the NCT.

The Census Bureau's working MENA classification includes countries and territories that were in the majority of MENA classifications used by other organizations in the United States, including state and federal government agencies, research organizations, and universities who classify countries and territories from the Middle East or North Africa. For the 2015 NCT, the Census Bureau classifies a person as MENA if they have ethnic origins or descent, roots, or heritage from any of the following nineteen countries (Algeria, Bahrain, Egypt, Iraq, Iran, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates, and Yemen) as well as the following ethnicities (Amazigh or Berber, Arab, Assyrian, Bedouin, Chaldean, Copt, Druze, Kurdish, and Syriac).

Wording of Instructions Dimension and Question Terminology Dimension. For these dimensions, we examine ways to improve the wording of question instructions, and whether alternative terminology or even no terms at all, help to improve the questions.

First, we focus on the ***wording of instructions***. The 2015 NCT will evaluate the use of different approaches for wording the instructions used to collect data on race and ethnicity. The 2010 AQE research found that respondents frequently overlook the instruction to "*Mark* [***X***] *one or more boxes*" and have difficulty understanding the instructions. From the 2010 AQE qualitative research we learned that some respondents stop reading the instruction after noticing the visual cue [***X***] and proceed directly to do just that – mark a box – overlooking the remainder of the instruction. The new instruction being tested in the 2015 NCT ("*Mark all boxes that apply*") is an attempt to improve the clarity of the question and make it more apparent that more than one group may be selected. The following options are being tested for this dimension in the 2015 NCT:

1	***"Mark [X] one or more"***	One version (old instructions) will advise respondents to, "*Mark* [***X***] *one or more boxes* ***AND*** *print* [origins/ethnicities/details]." *Note: Refer to Appendix A, panel #s 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, and 35*
2	***"Mark all that apply"***	An alternative version (new instructions), will advise respondents to, "*Mark all boxes that apply* ***AND*** *print* [origins/ethnicities/details] *in the spaces below. Note, you may report more than one group*." *Note: Refer to Appendix A, panel #s 2, 4, 6,8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, and 36*

Additionally, we focus on the dimension regarding ***question terminology***.

1	***"Race"* and *"Origin"* terms**	The use of "race" and "origin" as terminology (old instructions) will be used to guide respondents to answer the question (e.g., "*What is Person 1's race or origin?*"). *Note: Refer to Appendix A, panel #s 1-2, 7-8, 13-14, 19-20, 25-26, and 31-32*
2	***"Race"* and *"Ethnicity"* terms**	One alternative option being explored tests the use of both the terms "ethnicity" along with "race" in the question stem and/or instructions (e.g., "*What is Person 1's race or ethnicity?*")." *Note: Refer to Appendix A, panel #s 3-4, 9-10, 15-16, 21-22, 27-28, and 33-34*
3	**No terms at all ("*categories*")**	A second alternative option being explored tests the removal of the terms "race," "origin," and "ethnicity" from the question stem and instructions. Instead, a general approach asks, "*Which categories describe Person 1?*" *Note: Refer to Appendix A, panel #s 5-6, 11-12, 17-18, 23-24, 29-30, and 35-36*

<u>Instructions for AIAN Write-In Area</u>

The 2015 NCT will also examine different instructions to optimize detailed reporting within the AIAN write-in area. From the 2010 AQE research and recent 2014 qualitative research that the Census Bureau conducted with American Indians, Alaska Natives, and Central and South American Indian respondents, we know the instruction to "Print enrolled or principal tribe" causes confusion for many AIAN respondents and means different things to different people. The research found that AIAN respondents were confused by the use of different terms and concepts (e.g., "enrolled", "affiliated," "villages," "race," "origin," "tribe," etc.) and there was disagreement among focus group participants as to what "affiliated tribe" or "enrolled" or "villages" meant.

The overwhelming sentiment from 2014 AIAN focus group participants was that they want to be treated equally with other race/ethnic groups, and this was accomplished by not using different terminology (i.e., enrolled, affiliated, villages, etc.). The instruction "*Print, for example,...*" (along with AIAN example groups) allowed the respondents to understand what the question asked them to report and did not limit their write-in response by confounding the instructions with terms that mean different things to different people (e.g., tribes, villages, etc.). This instruction presented a viable alternative for further exploration in 2015 NCT research. Based on the findings and recommendations from this research, the 2015 NCT will test variations of the instructions for the AIAN write-in area to see how they perform.

1	***"Print enrolled or principal tribe, for example..."***	We plan to test the instruction, "*Print enrolled or principal tribe, for example...*" on control versions." *Note: Refer to Appendix A, panel #s 1, 13, 25*
2	***"Print, for example..."***	We plan to test the instruction, "*Print, for example...*" on experimental versions. *Note: Refer to Appendix A, panel #s 2-12, 14-24, 26-36*

Evaluating Performance of the Questions with New Web-Based Designs. As discussed above with the Question Format dimension, the 2015 NCT also presents the critical opportunity to compare the success of different question designs to determine how they perform in paper-based designs as well as with web-based data collection methods using the Internet, smartphone, and telephone response options. With the advantage of new technology to collect data via web-based designs, we are testing different versions of the Internet question with detailed checkboxes for soliciting detailed racial and ethnic origins, described above in Part A about Question Format.

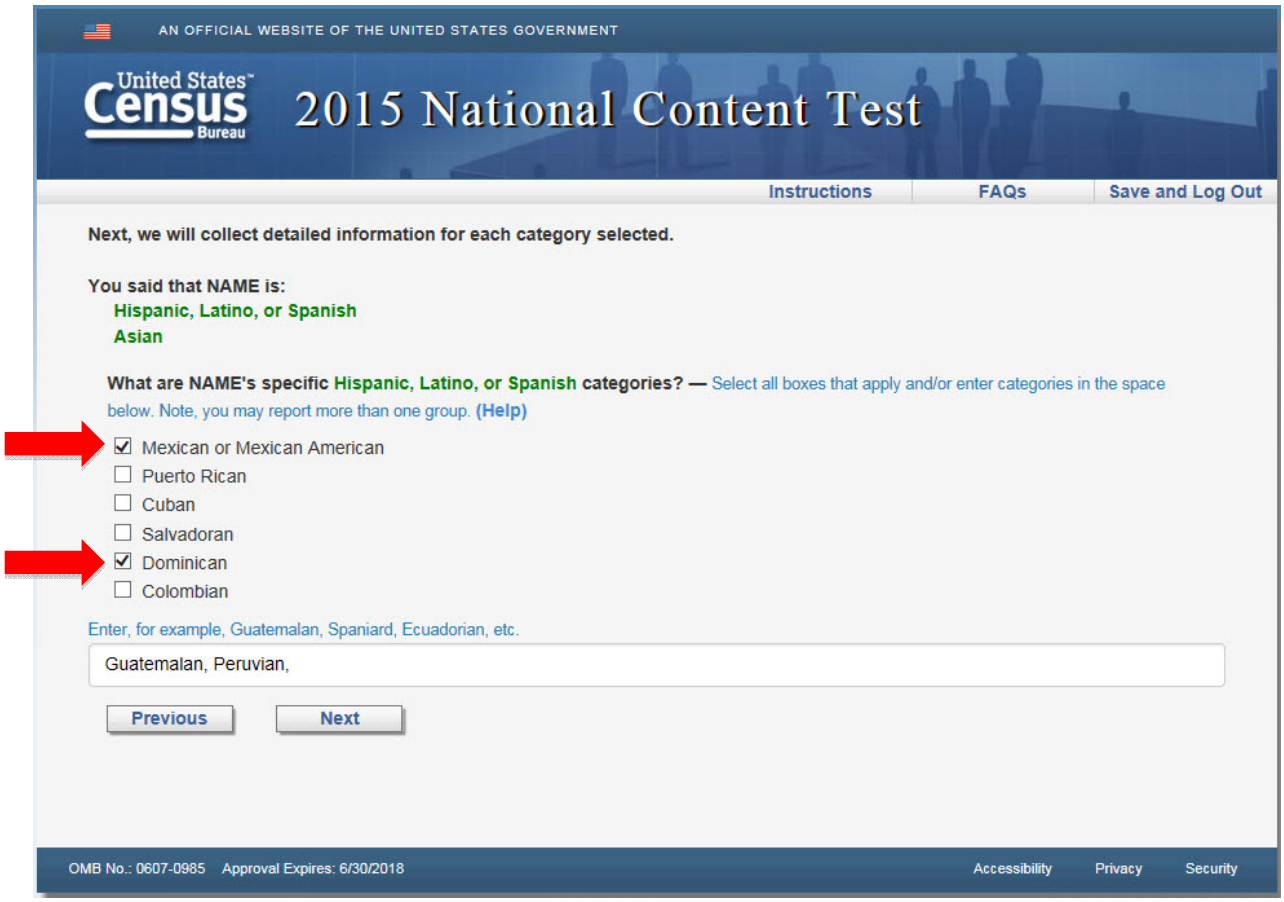
We expound upon this dimension below, to illustrate how this operates in the 2015 NCT. In the research, we employ these designs for all web-based methods, as we are not just limited to computer-based Internet responses. The 2015 NCT will enable people to answer via smartphone, land-line telephone, as well as with Telephone Questionnaire Assistance (TQA) from Census TQA representatives. The web-based approaches provide a series of screens to collect data for major groups (such as White, Hispanic, Black, and Asian) as well as data for detailed groups (such as Samoan, Iranian, Filipino, Jamaican, Puerto Rican, Irish, etc.). On the initial screen, we collect data on the major race/ethnic categories via a checkbox and examples, which are shown for the six largest detailed groups representing the geographic diversity of the OMB race/ethnic group's definition. Figure 6 provides an example where a respondent selects the box for Hispanic, Latino, or Spanish and the box for Asian.

Figure 6. Initial Screen for Web-Based Designs

After that, they select "NEXT" which will advance them to the next screen. For any selected category, a subsequent screen presents either several detailed checkbox groups and/or a dedicated write-in area to collect additional detailed responses, depending on the design treatment.

In our example, where the respondent marked they are Hispanic and Asian, the first follow-up screen collects detailed Hispanic groups, such as Mexican or Mexican American and Dominican (see Figure 7 below). Additionally, respondents can enter multiple additional responses, such as Guatemalan and Peruvian.

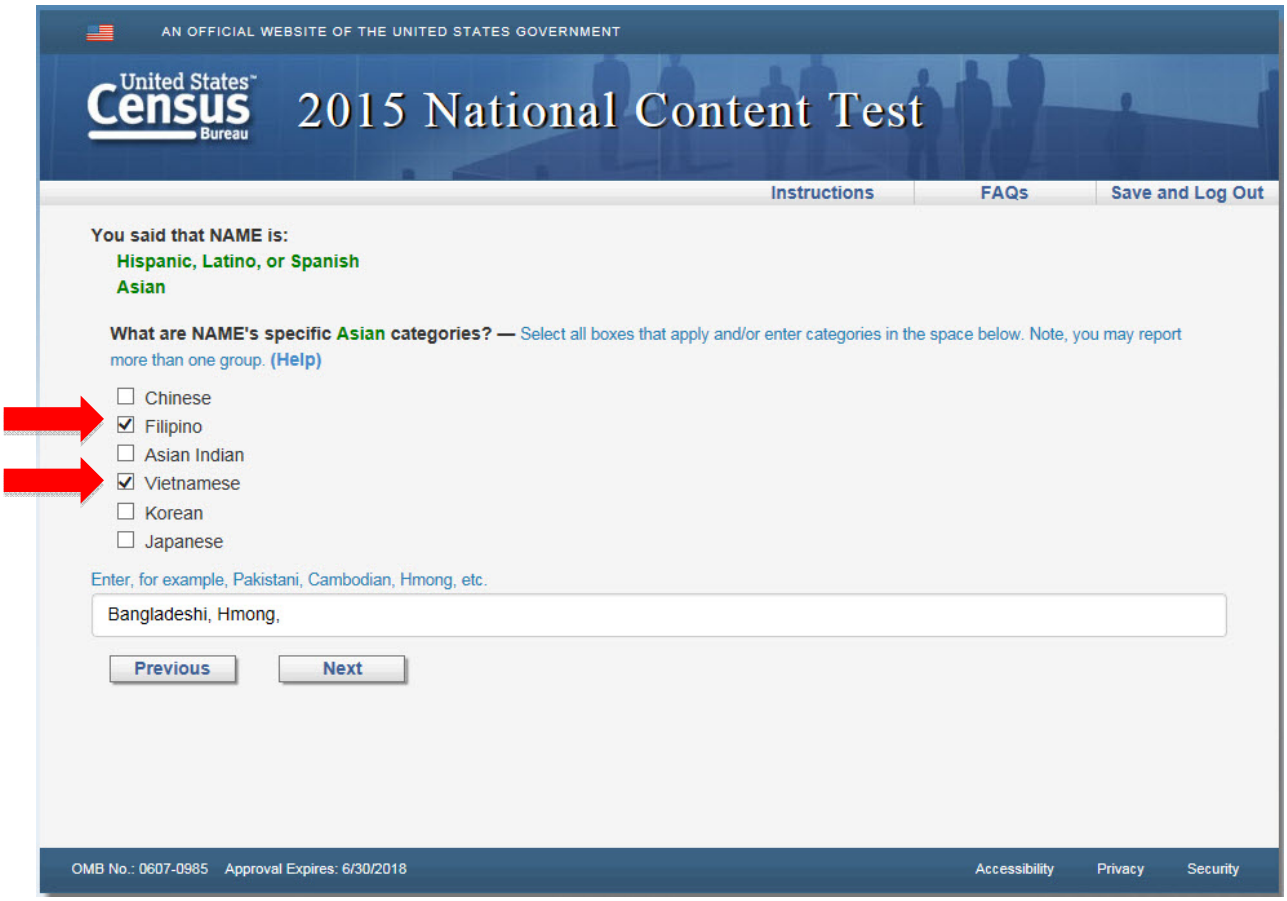
Figure 7. Subsequent Screen for Web-Based Designs (Hispanic, Latino, or Spanish)



After that, they would advance to the NEXT screen, where, in similar fashion, another screen collects detailed Asian responses, such as Filipino and Vietnamese (see Figure 8 below). Additionally, on this screen, respondents can enter multiple additional responses, such as Bangladeshi and Hmong.

Please note: Similar screens will collect detailed data for all communities, such as German, Jamaican, Lebanese, Samoan, etc.

Figure 8. Subsequent Screen for Web-Based Designs (Asian)



3.2 | Overview of 2015 National Content Test

The primary objective of the 2015 NCT is to test the content of the questionnaires. The content tested includes race/ethnicity, relationship, and coverage. These content items are tested by asking questions on these topics in several different ways. There are eight different versions of the stateside paper questionnaire, and two versions of the Puerto Rico questionnaire. There is more flexibility with testing on the Internet, allowing us to test 288 unique combinations of the different versions of questions to optimize self-response. In addition to testing content, the NCT is testing different contact strategies. Table 1 below displays the nine different contact strategies.

Table 1. Contact Strategy Panel Design

	Panel	#1 (August 24)	#2 (August 31)	#3* (September 8)	#4* (September 15)	#5* (September 22)
1	**Internet Push (Control)**	Letter	Postcard	Postcard	Mail Questionnaire	
2	**Internet Push with Early Postcard**	Letter	Postcard (August 25)	Postcard	Mail Questionnaire	
3	**Internet Push with Early Questionnaire**	Letter	Postcard	Mail Questionnaire	Postcard	
4	**Internet Push with Even Earlier Questionnaire**	Letter	Mail Questionnaire	Postcard	Postcard	
5	**Internet Choice**	Mail Questionnaire	Postcard	Postcard	Mail Questionnaire	
6	**Internet Push with Postcard as 3rd Reminder**	Letter	Postcard	Postcard	Mail Questionnaire	Postcard
7	**Internet Push Postcard**	Postcard	Postcard	Letter	Mail Questionnaire	
8	**Internet Push with Early Postcard and 2nd Letter Instead of Mail Q**	Letter	Postcard (August 25)	Postcard	Letter	
9	**Internet Push with Postcard and Email as 1st Reminder (Same time)**	Letter	Postcard and Email (August 25)	Postcard	Mail Questionnaire	

* Note: Targeted only to nonrespondents.

One final aspect that the NCT is testing is the language used in the letter. The first version of the letter is in English and has a Spanish sentence that encourages the Spanish-speaking respondent to reply using the Internet or TQA. The second version is a dual-sided letter, with English on one side Spanish on the other. Finally, the third version is a swim-lane letter, where the English text is on the left side of the page, and the Spanish text is on the right side of the page.

3.3 | Sample Design

Census Bureau researchers developed a nationally representative sample for the 2015 NCT. Overall, the sample included 1.2 million housing units in the United States, with a complex sampling design strategy for race and ethnicity. The 2015 NCT sampling methodology was designed to measure content testing differences for relatively small population groups. This sample design consisted of selecting 1,180,000 households from the fifty states and the District of Columbia and 20,000 households from Puerto Rico. The sampling frame was built from the Master Address File and excluded households that were selected in the 2015 American Community Survey and its supplements, were in the Savannah designated marketing area, were selected in the 2015 Census Test in Maricopa County, or had bad address values. Group quarters were also excluded.

The stateside sample design utilized a stratified, systematic sampling method that oversampled Census tracts that were susceptible to coverage overcounts, contained relatively high percentages of various race and ethnic groups, and had low self-response propensities. The stateside sample of 1,180,000 households was divided into three portions: coverage, race/ethnicity, and optimizing self response. To sample the coverage portion, the stateside sampling universe was subset to only include the tracts that had been flagged as susceptible for coverage overcounts. These subset tracts were stratified into the six special coverage groups, and then a sample of 180,000 households was selected for the coverage portion of the sample.

Next, the remaining households in the universe that were not selected for the coverage portion of the sample were stratified into one of the following six race strata, based on race, ancestry, and Hispanic origin data from 2010 Census data and 2009-2013 ACS data. The sample was designed to ensure that the unbiased estimates from the test accurately reflected the nation as a whole, across a variety of demographic characteristics, by oversampling various race and ethnic groups, including Asian and Native Hawaiian or Other Pacific Islander populations, American Indian or Alaska Native populations, Black or African American populations, Hispanic or Latino populations, and Middle Eastern or North African populations. The selection eligibility was done sequentially:

- **Middle Eastern or North African (MENA) Stratum**: Tracts where the percentage of people in the tract who identify as MENA was 10 percent or more.

- **American Indian or Alaska Native (AIAN) Stratum**: Tracts where the percentage of people in the tract who identify as AIAN was 10 percent or more.

- **Asian / Native Hawaiian or Other Pacific Islander (NHPI) Stratum**: Tracts where the percentage of people in the tract who identify as Asian or NHPI was 15 percent or more.

- **Black or African American Stratum**: Tracts where the percentage of people in the tract who identify as Black or African American was 25 percent or more.

- **Hispanic or Latino Stratum**: Tracts where the percentage of people in the tract who identify as Hispanic or Latino was 45 percent or more.

- **All Other Groups Stratum**: The remaining tracts that do not fall into one of the previous stratum.

The use of a multi-strata oversample of key population groups enables us to gain a broader and deeper universe of diverse population groups across the country. The map in Appendix E illustrates how the 2015 NCT sample is allocated for race and ethnicity. This research provides critical data to evaluate the key research questions, results, and findings of the 2015 NCT, which will inform recommendations to the OMB on the collection and classification of race and ethnicity data for the 2020 Census.

Table 2 below displays the estimated number of tracts, housing units, population, and race/ethnicity percentages in each of the six strata using ACS data. The table also displays the overall 2010 mail response rates for those tracts, which came from the Planning Database (PDB).

Table 2. Estimates of Race/Ethnicity Strata in the 2015 NCT Sample

Stratum	Number of Tracts	Number of HUs	Population	MENA %	AIAN %	Asian/ NHPI %	Black %	Hispanic %	2010 Mail Response Rate
MENA	770	1,176,267	3,112,719	**19.1**	1.0	14.2	9.4	15.5	67.4
AI**AN**	1,426	1,874,151	5,194,351	0.3	**26.3**	2.9	6.9	12.7	57.8
Asian/NH**PI**	6,359	10,555,098	30,048,814	2.2	1.3	**30.4**	9.0	20.0	69.0
Black	12,686	17,997,061	49,423,427	0.5	1.1	2.6	**53.5**	16.1	59.6
Hispanic	6,514	9,393,227	31,049,979	0.6	1.4	3.3	7.1	**73.6**	62.1
All Other	45,369	75,865,280	195,988,644	0.9	1.2	3.4	5.7	8.9	68.2

Source: Mathews (2015)

Table 3 below shows how the 800,000 housing unit NCT sample was allocated amongst the six strata.

Table 3. 2015 NCT Race/Ethnicity Sample Allocation

Stratum	Sample Size Chosen for Race Portion	Probability of Selection
MENA	100,000	.0767
AIAN	100,000	.0501
Asian/NHPI	100,000	.0086
Black	160,000	.0077
Hispanic	160,000	.0169
All Other	180,000	.0021

Finally, after the coverage and race/ethnicity portions had been selected, the remaining households in the universe were stratified into the three response propensity strata, and a sample of 200,000 housing units were selected for the OSR portion of the sample. After the 1.18 million household sample was selected, the sampling flags were assigned to indicate which version of the paper and Internet questionnaires the household would receive as well as which contact strategy would be used for that household. The use of a multi-strata oversample of key population groups will enable us to gain a broader and deeper universe of diverse population groups across the country.

Note, for the Puerto Rico sample, only the housing units in the San Juan municipio, or county, were eligible for the 2015 NCT. A systematic sample of 20,000 households was selected and the appropriate sampling flags were assigned. For more information on the sample design of the stateside or Puerto Rico sample, please refer to Mathews (2015).

3.4 | Reinterview Evaluation

3.4.1 | Overview of Reinterview

The second major component of the 2015 NCT will be a telephone reinterview study conducted with a sample of the 2015 NCT survey respondents. This reinterview research aimed to assess the accuracy and the reliability of the various race and Hispanic origin question designs by exploring responses to a number of probing questions. This research, which builds upon the successful reinterview that was conducted to the 2010 AQE research, will help to measure respondents' self-identified "true" racial and/or ethnic identity through a series of detailed questions and probes to aid in determining our truth measure. The 2015 NCT reinterview was developed in concert with demographic and sociological experts along with survey methodologists to probe more extensively into how respondents self-identify.

The 2015 NCT research includes a reinterview to assess the accuracy and reliability of the question alternatives for race and ethnicity. Approximately 100,000 households will be sampled for the reinterview, with approximately 75,000 of the cases focusing specifically on the topics of race and ethnicity. The remaining 25,000 households will be reinterviewed to study within-household coverage accuracy.

The telephone reinterview collects data from a subset of respondents and is focused solely on the race and Hispanic origin questions. The purpose of the reinterview questions is to ascertain the respondents' "true" self-identified racial and ethnic identities. We recognize that race and ethnicity are not quantifiable values. Rather, identity is a complex mix of one's family and social environment, historical or socio-political constructs, personal experience, context, and many other immeasurable factors.

Because this idea of "truth" is inherently difficult to define for self-identified race and Hispanic origin, we cannot expect to evaluate it with information from the self-response survey. However, we were able to employ an extensive series of detailed questions and probes to aid in determining our "truth" measure for the reinterview. This was a substantial addition to the AQE focus group research, and it yielded important results to help understand the data that were collected in the mail survey, as well as connections to the findings in the focus groups research.

The reinterview is designed to probe more extensively than the census questionnaire by asking three series of questions about how respondents self-identify, as well as collect more detailed information about respondents' racial and ethnic background. The first question is an open-ended question that asked the respondent to identify their race or origin. The second set of questions is a series of yes/no questions meant to probe into the respondent's complete racial and ethnic background. The third set of questions ask for detailed origin for each category that the respondent answered yes to. Later in the interview, there is also a question that asks how the respondent is perceived by others, but this question will not be used to determine truth. The selected question series underwent extensive cognitive testing, and were based on the 2010 AQE Reinterview questions (Dusch 2011). A copy of the race, ethnicity, and origin reinterview questions for the 2015 NCT are provided in Appendix C.

The data obtained in the reinterview will be used to estimate and compare two statistics important to a well-defined and stable measurement process. The first statistic is the bias in estimates of group membership that may occur if the responses are not an accurate reflection of the “true” self-identified group membership status due to imperfections in the design of the paper questionnaire. This statistic is called the “response” bias. The second statistic is the reliability of the measurement process over repeated measures (i.e., the paper questionnaire response and the reinterview response). This statistic is called the gross difference rate (GDR), which is used as a proxy for simple response variance.

Reinterview metrics include consistency scores, net difference rates ("accuracy", which provides a measure of over- or under-reporting for a category in the initial survey response), and gross difference rates ("reliability", defined as the percentage of responses for each group that were different from the initial response compared to the reinterview). Other key measures include item nonresponse rates, distributions (are changes to the question(s) associated with changes to the race/ethnicity distributions for each major group), and detailed reporting for each group. This will enable us to evaluate the key 2015 NCT research questions and results, and analyze the findings to inform recommendations to OMB on the collection and classification of race/ethnicity data for the 2020 Census and other federal surveys.

3.4.2 | Reinterview Sample Design

The reinterview sample design was a systematic random sub-sample of about one in eight (1:8) of the selected sample housing units. There was no further oversampling of the preselected reinterview cases beyond the oversampling present in the mailout sample. Those households for which we received a sufficient response before completion of the reinterview will be included in the reinterview workload.

3.5 | Data Processing

3.5.1 | Coding

As part of the 2015 NCT research, we will be collecting detailed data for all major categories via dedicated write-in areas and/or detailed checkboxes. This effort to collect, code, and tabulate myriad detailed groups required extensive research by Population Division’s Special Population Statistics Area (SPS) and consultation with external experts on various race and ethnic groups. Over the past two years, SPS subject matter experts researched and developed updates to enhance and expand previous code lists for Race, Hispanic Origin, and Ancestry data.

This research primarily focused on the code lists for European national origins and ethnic groups, Middle Eastern and North African national origins and ethnic groups, Sub-Saharan African national origin and ethnic groups, and Afro-Caribbean national origins and ethnic groups – many of which did not exist as distinct race codes in previous census data collections and tabulations – but also covered all major categories (e.g., Hispanic or Latino, Asian, Pacific Islander) within the full code range. These new code ranges, which include approximately 300 new terms, cumulatively create the 2015 Race, Ethnicity, and Origin Code List, which will be employed for the 2015 NCT research to help evaluate the reporting of detailed responses to the race/ethnic question(s).

For the 2015 NCT research, we employ the *Race, Ethnicity, and Origin Code List* to code race, ethnicity, and origin responses. The *Race, Ethnicity, and Origin Code List* employs 3-digit codes for race and ethnic groups in numeric and alphanumeric sequence, meaning that the 3-digit codes either contain all numbers (i.e., 114) or contain a mix of letters and numbers (i.e., A01). This critical reference is the foundational resource for researching and assigning codes to residual responses.

The list contains race, ethnicity, and origin codes and their unique descriptors, underneath larger headings for the major race/ethnic categories. Larger code ranges are used to group together and distinguish the major groups from one another (e.g., the 400-499 code range is reserved for Asian codes). The figure below lists the different code ranges, which represent the major race/ethnic groups, as well as a range for responses that do not fall within these groupings (i.e., United States, American, Religious responses, and Uncodable terms).

Figure 9. Range Values for Race, Ethnicity, and Origin Code List

Code Range Values	Race, Ethnicity, or Origin – General Headings
001-141, 182-194	White
200-299	Hispanic, Latino, or Spanish
300-399	Black or African American
400-499	Asian
500-599, A01-Z99	American Indian, Alaska Native, Central and South American Indian
142-181, 195	Middle Eastern or North African
600-699	Native Hawaiian or Other Pacific Islander
700-799	Some other race, ethnicity, or origin
800-899	Not used by coders
900-999	United States, American, Religious responses, Uncodable terms

3.5.2 | Data Editing

The race write-in response data from the test and the reinterview will be coded and pre-edited prior to data analysis. After the race and ethnicity responses have been coded (as explained in 3.5.1) the data collected from both the self-response and reinterview components of the NCT will be pre-edited using an updated and concise version of the 2010 Census procedures. The purpose of these pre-edits is to standardize the race and ethnicity classifications across all experimental panels.

The NCT pre-editing procedures will include:

- Converting checkbox responses to three-digit codes;
- Ensuring that codes assigned to write-in responses during the coding operation are valid;
- Limiting write-in responses to no more than 10 codes each;
- Eliminating duplicate codes; and
- Removing general codes when specific codes are provided (e.g., if the code for the Asian checkbox and a code for Chinese are present, only the code for Chinese will be retained).

3.5.3 | Name Matching

After the 2015 NCT reinterview data are prepared for analysis, the persons from the completed reinterview cases will be matched to the 2015 NCT persons within corresponding households.
The name matching process will use a computerized matching program developed to accurately link the correct person data from the two data collections together for analysis.

3.5.4 | Reinterview "Truth"

The NCT race and ethnicity reinterview includes a series of questions intended to explore the respondents' racial and ethnic background. Each respondent's "true" self-identified racial and ethnic identity will be determined through a combination of responses provided during the reinterview.
The following major categories (and multiple-group combinations of these categories) will be used for comparative purposes in the analyses:

1. White
2. Hispanic or Latino (Hispanic)
3. Black or African American (Black)
4. Asian
5. American Indian or Alaska Native (AIAN)
6. Middle Eastern or North African (MENA)
7. Native Hawaiian or Other Pacific Islander (NHPI)
8. Some Other Race, Ethnicity, or Origin (SOR)

The initial step of determining a respondent's "true" self-identified race/ethnicity will be through an automated match program, which will be applied to the coded reinterview responses. The first and second parts of the reinterview will be coded independently. Responses will be determined as "truth" for cases where both sets of questions had the same response, or where only one response was provided. Additional cases that are not auto-matched will be sent to clerical matching where analysts and demographers will independently study these special cases and decide on a final "truth" for each situation. Otherwise, unresolved cases will be sent to a panel review consisting of additional Census Bureau experts for final resolution. At all steps of the process, OMB standards will be used to guide final decisions on "truth."

3.6 | Variance Estimation

To account for the complex sample design of the experiment, we plan to use stratified jackknife replication estimation. Due to software and processing limitations, we plan to use a random groups method to create the replicates. In this method, housing units are sorted in the order they were selected and reassigned to one of 250 different groups, or replicates. To help ensure the validity of statistical inference when making multiple statistical comparisons, when applicable, multiple comparison corrections will be used to maintain the family-wise error rate at $\alpha = 0.05$. Dunn's procedure will be performed to adjust for the increased possibility of erroneous conclusions when multiple comparisons adjustment procedures are used. The multiple comparison corrections reduce the possibility of identifying false-positive differences and ensure that we do not cloud our ability to form inferential conclusions.

3.7 | Introduction to Data Analysis and Decision Criteria for Content Decisions

As previously discussed, the 2015 NCT is the primary mid-decade opportunity to compare different content prior to making final decisions about the content for the 2020 Census. The test will include a reinterview to assess the accuracy and reliability of the question alternatives for race and Hispanic origin, which will enable us to evaluate the key research questions, results, and findings to inform recommendations for the 2020 Census content on race and ethnicity. As discussed previously, data from the NCT reinterview is critical for the analysis in determining consistency between the self-response survey and respondents' "truth." The 2015 NCT panels are designed to test key dimensions for improving the questions on race and ethnicity. These dimensions are expected to impact different aspects of respondent reporting, including item reliability/consistency, nonresponse, and distributions. All of these aspects must be considered when making recommendations on the various dimensions and overall question design.

The inherent complexity of this experiment and its many complex analyses require that decision criteria be developed to assist with making recommendations on each of the key dimensions being explored and choosing the "best" overall question design for the 2020 Census. These decision criteria provide important statistical measures and hypotheses for each measure, which provide indications of favorable and unfavorable results. An interdivisional group will convene to determine the success or failure of the key dimensions and determining the best overall panel based on these and other specific criteria.

This section presents the research questions we will be exploring for each of the dimensions of the study. The questions are listed in order of priority for determining recommendations based on the described decision criteria. Each research question is accompanied by a brief explanation of what is being investigated, along with table shells for some of the response distributions and other analyses that

will be conducted to compare and evaluate the various race and ethnicity treatments in the 2015 NCT. Each of the analyses will be produced by mode (Internet, phone, paper), though the overall response estimates will be the focus of the report. Finally, a descriptive statement is provided to illustrate the decision criteria.

3.7.1 | Analysis Measures

The data analysis will be performed for the 2015 NCT response data and the 2015 NCT reinterview response data. Some of the major analysis measures that will be employed for this research are listed below:

- Major race/ethnic category response distributions.
- Item nonresponse rates.
- Net difference rates (as a measure of bias), based on content reinterview.
- Percentage of consistent responses (measure of reliability) for each major race/ethnicity group, based on content reinterview.
- Percentage of detailed responses for each major race/ethnicity group.
- Percentage of multiple responses for major groups, compared with the reinterview.

In addition, the analysis will incorporate multivariate regression models, as appropriate, to determine which dimensions of the race and ethnicity research are most associated with differences for each of the statistical metrics of interest. The regression models will consider factors such as the experimental treatment, device type (e.g., computer, tablet, or smartphone), mode of response, and demographic characteristics.

3.8 | Testing Alternative Question Formats (Separate vs. Combined)

Objective: To Determine the Best Question Format for Collecting Race/Ethnic Data

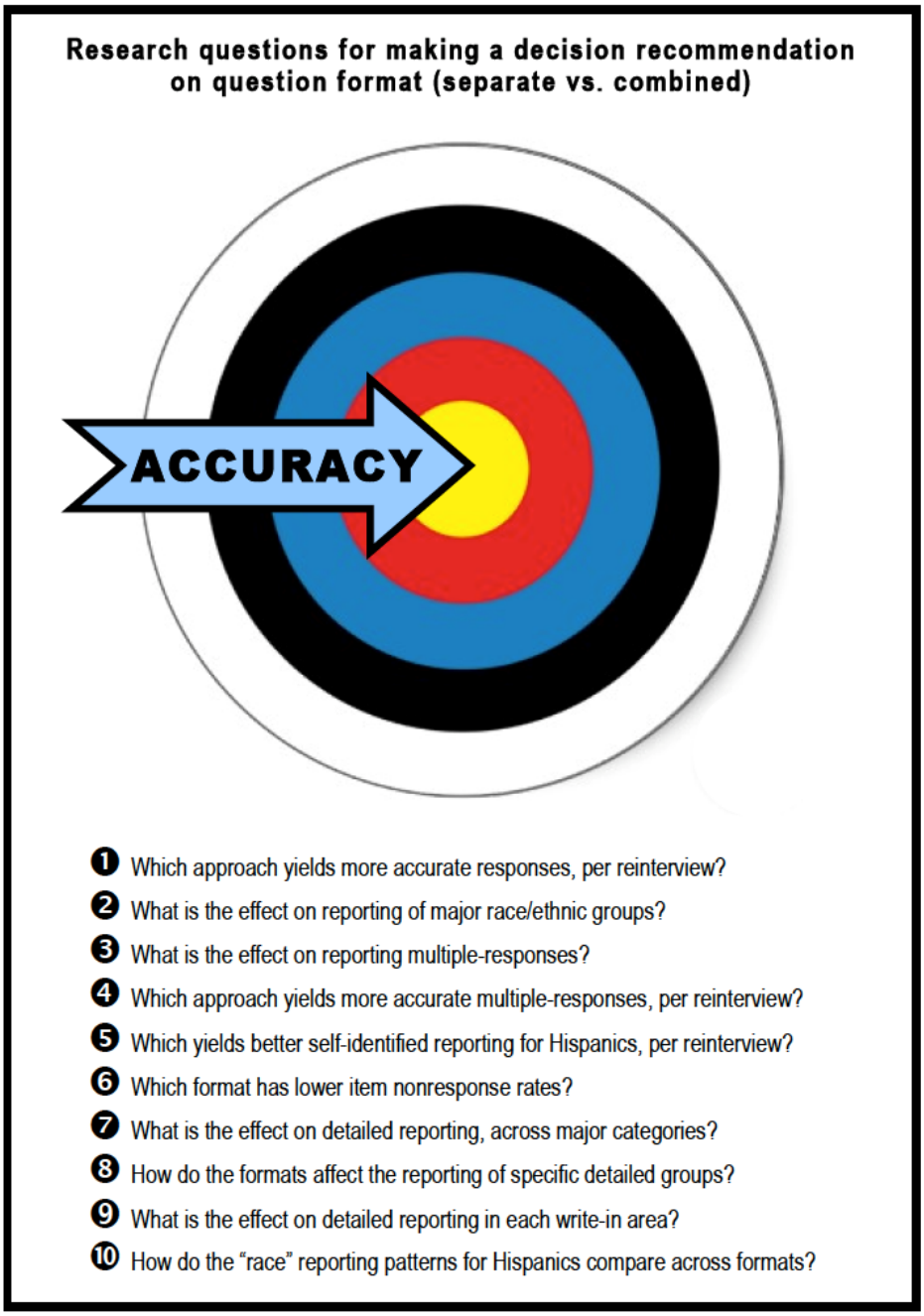
Goal: Maintain or improve the quality of the race and Hispanic origin data by using a combined race and ethnicity question.

One main objective of the 2015 NCT is to evaluate the use of different question format approaches for collecting data on race and ethnicity. This objective builds upon the successful findings of the 2010 AQE research which showed promising strategies for addressing the challenges and complexities of race and ethnicity measurement and reporting. The AQE research led to experimental question refinements for the design of the 2015 NCT question formats, which are being tested in various ways.

One approach uses two separate questions – the first about Hispanic origin and the second about race. The other approach combines the two separate questions into one question about race and ethnicity. The combined question approach also tests different designs for collecting detailed responses – one design employs dedicated write-in areas to collect detailed responses, the other design uses detailed checkboxes and a subsequent write-in area to collect detailed responses.

The decision criteria for the Question Format Dimension focus on evaluating the different approaches for the presentation of the race/ethnic question (or questions). Specific analyses for this dimension will examine the research questions outlined in this section, which will be used to tease out benefits and drawbacks of the various treatments. A recommendation on the best question format for collecting and producing data on race/ethnicity for the 2020 Census (separate vs. combined with write-in areas vs. combined with detailed checkboxes) will be made based on results from the 2015 NCT. An overview of the research questions that we are focusing on for this dimension are shown in the figure below.

Figure 10.



The overarching principle of this research is to find ways to improve the ***accuracy*** of data on race and ethnicity. Accuracy is the most important goal, and the research questions that we are examining yield insights to different aspects of accuracy. The ten research questions outlined in this section will explore different aspects of the ways in which we seek to improve the accuracy of data on race and ethnicity. We note that all of the research questions are important, as they help us to understand the accuracy that we are measuring – whether it be the accuracy of reporting for major race/ethnic groups (such as White, Hispanic, Black, Asian, etc.), the accuracy of reporting multiple responses (such as White ***and*** Black, Black ***and*** American Indian, White ***and*** Asian ***and*** Pacific Islander, etc.), or the accuracy of reporting detailed nationalities or ethnic groups (such as Irish, Puerto Rican, Jamaican, Filipino, etc.).

Within our analyses of the different research questions, the decisions regarding the dimension of question format predominantly rely on the following major factors: (a) better measures of "truth;" (b) yielding self-identified reporting within OMB categories; (c) levels of detailed reporting; and (d) lower item nonresponse. The foundation of determining which question format is best rests on measuring "truth."

As described previously in Section 3.4, the reinterview will enable us to measure respondents' self-identified "true" racial and/or ethnic identity through a series of detailed questions and probes, which are compared to their responses on the self-response survey. Ultimately, these explorations will help us evaluate which question format yields more accurate and reliable results to reflect respondents' self-identification.

Strategies for Evaluating Success

Our strategies for evaluating the success of the different question format approaches for collecting data on race and ethnicity focus on the following series of key research questions. In our analyses, we will be looking to see which of the questions formats works best (separate questions approach vs. combined question approach with write-in areas vs. combined question with detailed checkboxes). Within this, we will focus on the following factors:

- Which question format approach yields the greatest accuracy?
- Which question format approach enables respondents to fully self-identify and reflect their "truth"?
- Which question format optimizes the reporting of multiple races and multiple ethnic origins?
- Which question format improves reporting of detailed nationalities and ethnic origins?
- Which question format reduces item nonresponse?
- Which question format improves reporting in major OMB categories (reducing "Some Other Race" reporting)?

The next several pages present a thorough, detailed walkthrough of the research questions, data tables, and decision criteria for analyzing the results for question format. Our hypotheses are presented as research questions (a priori), along with table shells and explanations of what will be analyzed. We also present decision criteria to explain how the results will be evaluated in order to make recommendations on the research question that is posed.

Question 1. Which question approach (Separate vs. Combined) yields more accurate responses, per the reinterview?

Our analysis will compare the levels of reporting for each race/ethnic response category in the survey and with those reported in the 2015 NCT reinterview. We will examine the estimated percentages of people in each of the following major categories:

1. White
2. Hispanic or Latino (Hispanic)
3. Black or African American (Black)
4. Asian
5. American Indian or Alaska Native (AIAN)
6. Middle Eastern or North African (MENA)
7. Native Hawaiian or Other Pacific Islander (NHPI)
8. Some Other Race, Ethnicity, or Origin (SOR)
9. Multiple Responses

The reinterview data will provide measures of response bias and gross difference rates, which we will examine in conjunction with respondent's "true" self-identified racial and ethnic identity as determined through the reinterview. These data will enable us to evaluate the patterns of consistency for responses between the initial survey self-response and the reinterview responses for each of the major race/ethnic groups being studied.

For responses provided in the reinterview, we will compute the percentage that provided the same response category in the initial survey. This will be a critical factor for determining which format (Separate or Combined) yields more accurate responses reflecting respondents' "true" self-identified racial and ethnic identity.

Table 4. Percentage of Consistent Responses between Self-Response and Reinterview for Major Race/Ethnic Groups by Question Format

	White	Hispanic	Black	Asian	AIAN	MENA	NHPI	SOR	Multiple
Separate Question Approach	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Write-In Response Areas	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Detailed Checkboxes	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

In addition, we will compare the race and ethnicity distributions for the self-response and reinterview in order to compute net difference rates for each major group. This is a measure of response bias, and question formats that are more accurate will have a lower value. For example, if 55.0 percent of respondents identify in one group during the reinterview and 53.5 percent of reinterview respondents identify in that same group, then the net difference rate for the group is 1.5 percent.

Table 5. Reinterview Race Reporting Compared to Self-Response for Each Question Format

Separate Questions Approach

	Reinterview								
Self-Response	**White**	**Hispanic**	**Black**	**Asian**	**AIAN**	**MENA**	**NHPI**	**SOR**	**Multiple**
White	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
NHPI	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Multiple	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Combined Question with Write-In Response Areas

	Reinterview								
Self-Response	**White**	**Hispanic**	**Black**	**Asian**	**AIAN**	**MENA**	**NHPI**	**SOR**	**Multiple**
White	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
NHPI	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Multiple	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Combined Question with Detailed Checkboxes

	Reinterview								
Self-Response	**White**	**Hispanic**	**Black**	**Asian**	**AIAN**	**MENA**	**NHPI**	**SOR**	**Multiple**
White	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
NHPI	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Multiple	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Decision Criteria for Question 1: The tables for research question 1 will provide metrics on the consistency and accuracy of reporting between the initial self-response to the survey and the responses in the reinterview. We will examine the percentages of consistency for each of the major race/ethnic groups by question format, to determine which approach has greater consistency of responses.

We will examine how the bolded results along the diagonal inform us how consistent report is between the two measures (i.e., how close to 100% are the self-response and reinterview response for each major category?).

With these insights, results from the reinterview will determine which question format (Separate or Combined), produces the most reliable and accurate responses for reflecting respondents' "true" self-identified racial and ethnic identity.

❷ Question 2. What is the effect of the different formats (Separate vs. Combined) on the reporting of major racial and ethnic groups (e.g., White, Hispanic, Black, Asian, etc.) and the reporting of "Some other race or ethnicity?"

Our analysis will examine reporting patterns of major racial and ethnic groups, for the different question formats in the initial survey response. We will examine the estimated percentages of people in each of the following major categories:

- White
- Hispanic or Latino (Hispanic)
- Black or African American (Black)
- Asian
- American Indian or Alaska Native (AIAN)
- Middle Eastern or North African (MENA)
- Native Hawaiian or Other Pacific Islander (NHPI)
- Some Other Race, Ethnicity, or Origin (SOR)
- Multiple Responses
- Invalid Responses
- Missing Responses

The category for "Invalid Responses" represents cases when a respondent responded to the question but provided a response that was either coded as "invalid" or the reporting pattern was determined to be "invalid" based on reporting requirements. For example, some respondents in the past have reported that they are a "*Martian*" or a "*Human Being*," or they wrote in objections to the question such as, "*This is none of your business*" or "*We are all one human race*." Additionally, some respondents in the past have marked every single checkbox category but did not provide valid write-in responses, and these response patterns are edited and determined to be "invalid." The "Missing Responses" category represents cases when a respondent does not answer the question at all – they do not mark any of the checkboxes, nor do they enter a response in any of the write-in areas.

Table 6. Race/Ethnic Group Distribution by Treatment

	Separate Question	Combined Question with Write-In Response Areas	Combined Question with Detailed Checkboxes
White	% (se)	% (se)	% (se)
Hispanic	% (se)	% (se)	% (se)
Black	% (se)	% (se)	% (se)
Asian	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)
NHPI	% (se)	% (se)	% (se)
SOR	% (se)	% (se)	% (se)
Multiple	% (se)	% (se)	% (se)
*** Invalid**	% (se)	% (se)	% (se)
*** Missing**	% (se)	% (se)	% (se)

In addition, Table 7 enables us to examine these reporting patterns in finer detail regarding the particular race and Hispanic responses that are reported. We examine this in detail by evaluating the types of responses that are reported in the different question formats.

Table 7. Distribution of Hispanic Responses and Non-Hispanic Responses by Treatment

	Hispanic		Not Hispanic								*	*
	Hispanic alone	**Hispanic + Major Race Group**	**White alone**	**Black alone**	**Asian alone**	**AIAN alone**	**MENA alone**	**NHPI alone**	**SOR alone**	**Multiple Responses**	**Invalid**	**Missing**
Separate Question	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Write-In Response Areas	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Detailed Checkboxes	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

Decision Criteria for Question 2: We will evaluate results of the relative percentages for people in each of the major categories shown in the tables above. Examining the results in the tables, we expect:

- Similar percentages of Hispanic, Black, Asian, AIAN, MENA, and NHPI groups should be seen for the different formats.
- The percentage of people reporting in the SOR category should be lower in the Combined approach than in the Separate Question approach.
- The percentage of the White alone population in the Combined approach should be similar or lower than the percentage of the White alone, non-Hispanic population in the Separate Question approach.

❸ Question 3. What is the effect of the different formats (Separate vs. Combined) on the reporting of multiple-responses (e.g., people who report White *and* Black; or White *and* Hispanic; or White *and* Asian; etc.)?

We will examine the overall level of multiple-responses reported for the separate questions approach and for the combined question approach. We will also examine the level of reporting multiple-responses for each race/ethnic category (e.g., White in combination with one or more other groups).

Table 8. Reporting of Multiple-Responses in Different Question Formats

Level of multiple responses reported by CATEGORY	Separate Question	Combined Question with Write-In Response Areas	Combined Question with Detailed Checkboxes
WHITE in combination with another group	% (se)	% (se)	% (se)
HISPANIC in combination with another group	% (se)	% (se)	% (se)
BLACK in combination with another group	% (se)	% (se)	% (se)
ASIAN in combination with another group	% (se)	% (se)	% (se)
AIAN in combination with another group	% (se)	% (se)	% (se)
MENA in combination with another group	% (se)	% (se)	% (se)
NHPI in combination with another group	% (se)	% (se)	% (se)
SOR in combination with another group	% (se)	% (se)	% (se)

Decision Criteria for Question 3: Table 8 provides the levels of multiple-responses reported within each of the major race/ethnic categories. We will examine the overall level of multiple-responses reported for the separate questions approach and for the combined question approach to see which yield higher or lower reporting, and for which groups. We will also examine the level of reporting multiple-responses for each race/ethnic category (e.g., White in combination with one or more other groups). We will evaluate results of the relative percentages of multiple responses in each of the major categories, and expect to find that:

- Similar or higher percentages of multiple-group reporting should be seen in the combined question format for Black, Asian, AIAN, and NHPI groups, compared with the Separate approach.

- The percentage of people reporting SOR and another group should be the same or lower in the Combined approach than in the Separate Question approach. The percentage of multiple-group reporting for Hispanics will be higher in the Separate Questions approach. This is related to what we expect to find with respect to research question 1, as we evaluate the accuracy of reporting between the reinterview and self-response.

4

Question 4. Which question approach (Separate vs. Combined) yields more accurate multiple-response data (e.g., White *and* Black; White *and* Asian), per the reinterview?

We will examine the overall level of consistency for multiple-responses (as a total) between the survey and reinterview. Using the results in Table 9 below, we will compare results for the separate questions approach and the combined question approach.

Table 9. Overall Consistency of Self-Response and Reinterview for Multiple-Responses by Question Format

	Consistency of Multiple-Responses (Survey vs. Reinterview)
Separate Questions	% (se)
Combined Question with Write-In Response Areas	% (se)
Combined Question with Detailed Checkboxes	% (se)

In addition, we will examine this in finer detail by exploring the reporting of major multiple-response combination groups and comparing their percentages in the 2015 NCT reinterview. Based on results from 2010 Census and 2010 AQE, the expected major multiple-response categories for this analysis, as shown in Table 10 and Table 11, are listed below (we may also examine other multiple-response combinations greater than 1 percent).

1. White ***and*** Black
2. White ***and*** Hispanic
3. White ***and*** Asian
4. White ***and*** AIAN
5. Hispanic ***and*** Black

Table 10. Percentage of Consistent Responses between Self-Response and Reinterview for Selected Multiple-Response Groups by Question Format

	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black
Separate Question	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Write-In Response Areas	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Detailed Checkboxes	% (se)	% (se)	% (se)	% (se)	% (se)

Table 11. Self-Response Compared to Reinterview Response for Selected Multiple-Response Groups

Separate Questions Approach

	Reinterview						
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Other combinations	Single response
White *and* **Black**	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Hispanic**	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Asian**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White *and* **AIAN**	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
Hispanic *and* **Black**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Other combinations	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single response	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Combined Question with Write-In Response Areas

	Reinterview						
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Other combinations	Single response
White *and* **Black**	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Hispanic**	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Asian**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White *and* **AIAN**	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
Hispanic *and* **Black**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Other combinations	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single response	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Combined Question with Detailed Checkboxes

	Reinterview						
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Other combinations	Single response
White *and* **Black**	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Hispanic**	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Asian**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White *and* **AIAN**	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
Hispanic *and* **Black**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Other combinations	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single response	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Decision Criteria for Question 4: The tables for research question 4 will provide metrics on the consistency of reporting between the initial self-response to the survey and the responses in the reinterview. We will examine the percentages of consistency for each of the major multiple-response groups by question format, to determine which approach has greater consistency of responses.

We will examine how the bolded results along the diagonal inform us how consistent reporting is between the two measures (i.e., how close to 100% are the self-response and reinterview response for each category?). These results from comparing the consistency of responses among major multiple-response groups (e.g., White ***and*** Black; White ***and*** Hispanic; etc.) will determine which question format (Separate or Combined), yields the greatest consistency for reporting multiple-group responses, and reflecting the respondents' "true" self-identified racial and ethnic identity.

Question 5. Which question approach yields better self-identified reporting among Hispanics, reflecting lower "Some Other Race" reporting and more accurate responses based on the reinterview?

Exploring the results for research question 5 involves a complex comparison, which focuses on the types of response patterns for respondents who, from the reinterview data, have a "true" self-identified response as Hispanic, as well as those who self-identify as Hispanic and another group. Examining the results for these respondents, we examine their responses in the initial survey to determine whether the inclusion or exclusion of a Hispanic or Latino category enabled them to fully report their Hispanic identity.

In the separate question approach, we look to see whether Hispanics either:

1. answer the separate race question by reporting a response within an OMB race category (i.e., White; Black; etc.),
2. leave the separate race question unanswered, or
3. provide a "Hispanic" response to the separate race question, which is ultimately classified as "Some Other Race."

In the combined question approach, we look to see whether Hispanics:

1. also report a response within an OMB race category (i.e., White; Black; etc.), or
2. only report a "Hispanic" response
3. provide a "Hispanic" response in the "Some Other Race" response area.

As shown in Table 12, these response patterns will be evaluated in conjunction with the NCT reinterview data to determine which approach (Separate or Combined) yields better Hispanic "race" reporting (or non-reporting), reflecting lower "Some Other Race" reporting and more accurate responses, per the reinterview, for respondents of Hispanic origin.

Table 12. Reporting Patterns of the Hispanic Reinterview Population to the Survey Question Formats

	Separate Question	Combined Question with Write-In Response Areas	Combined Question with Detailed Checkboxes
Identified as Hispanic ONLY	% (se)	% (se)	% (se)
Identified as Hispanic *AND* White	% (se)	% (se)	% (se)
Identified as Hispanic *AND* Black	% (se)	% (se)	% (se)
Identified as Hispanic *AND* another group(s) (e.g., Asian, AIAN, etc.)	% (se)	% (se)	% (se)
Identified as NOT Hispanic	% (se)	% (se)	% (se)
Did not provide a response to survey	% (se)	% (se)	% (se)
Reported an invalid response	% (se)	% (se)	% (se)

Decision Criteria for Question 5: The results from the reinterview shown in Table 12 will help illuminate which question format enables Hispanic respondents to fully report their Hispanic identity. We expect to find that the combined questions will results in lower "Some Other Race" reporting and more accurate responses, per the reinterview, compared with the Separate Questions approach.

6 **Question 6. Which format (Separate vs. Combined) has lower item nonresponse rates?**

We will examine the overall level of item nonresponse for the separate questions approach and for the combined question approach. As discussed in the overview of the research dimensions (Section 3.1.2), the combined question format will have one item nonresponse rate for the race/ethnic question, whereas the separate question approach will have three different item nonresponse rates that can be analyzed: one for the Hispanic Origin question, one for the Race question, and that considers whether there is a response to either question. Table 13 will examine each of these results.

Table 13. Pattern of Response by Question Format

What percentage of respondents provide…?	No response to Hispanic Origin question	No response to Race question	No response to either question
Separate Question	% (se)	% (se)	% (se)
Combined Question with Write-In Response Areas	NA	% (se)	NA
Combined Question with Detailed Checkboxes	NA	% (se)	NA

Decision Criteria for Question 6:

Results for the race item nonresponse rate for the combined question should be the same or lower than the item nonresponse rate for the separate question approach.

Question 7. What is the effect of the different formats (Separate vs. Combined) on detailed group reporting across major categories (e.g., detailed reporting for Whites; detailed reporting for Hispanics; detailed reporting for Blacks; detailed reporting for Asians; etc.)?

We will examine the overall level of detailed group reporting for each of the major categories between the separate question formats and the combined question formats. For example, we will evaluate the percentage of detailed Hispanic responses obtained via the separate Hispanic origin question in comparison with the combined question. Similarly, we will evaluate the percentage of detailed Asian responses obtained via the separate race question in comparison with the combined question. Examining the results for every major group, we will evaluate which format (Separate vs. Combined with write-in areas vs. Combined with detailed checkboxes) maximizes detailed group reporting across all major categories.

Table 14. Detailed Reporting for Major Race/Ethnic Groups by Treatment
(Percentage providing detailed responses)

Alone or in Combination Groups	Separate Question	Combined Question with Write-In Response Areas	Combined Question with Detailed Checkboxes
White	% (se)	% (se)	% (se)
Hispanic*	% (se)	% (se)	% (se)
Black	% (se)	% (se)	% (se)
Asian*	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)
NHPI*	% (se)	% (se)	% (se)
SOR	% (se)	% (se)	% (se)

(Note, the rows for Hispanic, Asian, and NHPI are marked with an asterisk (*) because these are the only groups with dedicated detailed checkboxes in the Separate Questions format.)

Decision Criteria for Question 7: We will examine the results in Table 14 to evaluate the relative percentages of detailed reporting in each of the major categories:

- The level of detailed group reporting for all major categories (* other than Hispanic, Asian, and NHPI) should be equivalent in the Separate approach and the Combined approach with write-in areas.
- Detailed reporting in the Separate approach for Hispanics, Asians, and NHPIs (formats which include dedicated detailed checkboxes for these groups) should be compared with the Combined approach with detailed checkboxes (web-based designs) and the multiple-detailed checkboxes approach for the Combined question (paper format Panel W).
- If detailed reporting is higher in the combined approach for groups such as White, Black, AIAN, or MENA, this would indicate a favorable design.

Question 8. How do the different question formats (Separate vs. Combined) affect the reporting of detailed groups (e.g., dedicated detailed check boxes vs. only write-in areas)?

One of the key design differences among the different question formats is the employment of major category checkboxes in conjunction with either detailed group checkboxes or write-in areas to collect detailed responses. For the separate question approach, we will examine the level of detailed reporting for categories with detailed checkbox groups (i.e., Hispanics, Asians, and Pacific Islanders) in comparison to combined question approaches where detailed checkboxes are also provided for these groups.

We will also examine the level of detailed reporting for categories without detailed checkbox groups (i.e., Whites, Blacks, and AIANs) for the separate question approach and the combined question approach to evaluate how response rates compare for these groups.

Finally, we will compare the level of detailed reporting within the different combined question approaches, to see whether detailed checkbox formats (Panel W) perform better than formats which employ write-in response areas. For example, we will examine the percentage of White respondents who reported multiple White detailed groups (e.g., German and Irish), Hispanic respondents (e.g., Dominican and Puerto Rican), Pacific Islander respondents (e.g., Native Hawaiian and Samoan); etc.

Table 15. Reporting of Detailed Groups by Treatment

(Note, the data table below will provide tallies of responses for each category – e.g., a response of "German, Irish, African American" will be tallied on all three respective lines as a detailed response.)

	Separate Question (with Dedicated Checkboxes)		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
	Group has Dedicated Checkbox	Must Use Write-In Area	Group has Dedicated Checkbox	Must Use Write-In Area	Group has Dedicated Checkbox	Must Use Write-In Area
WHITE Detailed Groups:						
German	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Irish	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
English	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Italian	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Polish	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
French	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Additional "White" detailed responses	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
HISPANIC Detailed Groups:						
Mexican or Mexican American	**% (se)**	NA	NA	**% (se)**	**% (se)**	NA
Puerto Rican	**% (se)**	NA	NA	**% (se)**	**% (se)**	NA
Cuban	**% (se)**	NA	NA	**% (se)**	**% (se)**	NA
Salvadoran	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Dominican	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Colombian	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA
Additional "Hispanic" detailed responses	NA	**% (se)**	NA	**% (se)**	**% (se)**	NA

(Note: Table continues on next page)

	Separate Question (with Dedicated Checkboxes)		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
	Group has Dedicated Checkbox	**Must Use Write-In Area**	**Group has Dedicated Checkbox**	**Must Use Write-In Area**	**Group has Dedicated Checkbox**	**Must Use Write-In Area**
BLACK Detailed Groups:						
African American	NA	% (se)	NA	% (se)	% (se)	NA
Jamaican	NA	% (se)	NA	% (se)	% (se)	NA
Haitian	NA	% (se)	NA	% (se)	% (se)	NA
Nigerian	NA	% (se)	NA	% (se)	% (se)	NA
Ethiopian	NA	% (se)	NA	% (se)	% (se)	NA
Somali	NA	% (se)	NA	% (se)	% (se)	NA
Additional "Black" detailed responses	NA	% (se)	NA	% (se)	% (se)	NA
ASIAN Detailed Groups:						
Chinese	% (se)	NA	NA	% (se)	% (se)	NA
Filipino	% (se)	NA	NA	% (se)	% (se)	NA
Asian Indian	% (se)	NA	NA	% (se)	% (se)	NA
Vietnamese	% (se)	NA	NA	% (se)	% (se)	NA
Korean	% (se)	NA	NA	% (se)	% (se)	NA
Japanese	% (se)	NA	NA	% (se)	% (se)	NA
Additional "Asian" detailed responses	NA	% (se)	NA	% (se)	% (se)	NA
AIAN Detailed Groups:						
Navajo Nation	NA	% (se)	NA	% (se)	NA	% (se)
Blackfeet Tribe	NA	% (se)	NA	% (se)	NA	% (se)
Mayan	NA	% (se)	NA	% (se)	NA	% (se)
Aztec	NA	% (se)	NA	% (se)	NA	% (se)
Native Village of Barrow Inupiat Traditional Government	NA	% (se)	NA	% (se)	NA	% (se)
Nome Eskimo Community	NA	% (se)	NA	% (se)	NA	% (se)
Additional "AIAN" detailed responses	NA	% (se)	NA	% (se)	% (se)	NA
MENA Detailed Groups:						
Lebanese	NA	% (se)	NA	% (se)	NA	% (se)
Iranian	NA	% (se)	NA	% (se)	NA	% (se)
Egyptian	NA	% (se)	NA	% (se)	NA	% (se)
Syrian	NA	% (se)	NA	% (se)	NA	% (se)
Moroccan	NA	% (se)	NA	% (se)	NA	% (se)
Algerian	NA	% (se)	NA	% (se)	NA	% (se)
Additional "MENA" detailed responses	NA	% (se)	NA	% (se)	% (se)	NA

(Note: Table continues on next page)

	Separate Question (with Dedicated Checkboxes)		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
	Group has Dedicated Checkbox	Must Use Write-In Area	Group has Dedicated Checkbox	Must Use Write-In Area	Group has Dedicated Checkbox	Must Use Write-In Area
NHPI Detailed Groups:						
Native Hawaiian	% (se)	NA	NA	% (se)	% (se)	NA
Samoan	% (se)	NA	NA	% (se)	% (se)	NA
Chamorro	% (se)	NA	NA	% (se)	% (se)	NA
Tongan	NA	% (se)	NA	% (se)	% (se)	NA
Fijian	NA	% (se)	NA	% (se)	% (se)	NA
Marshallese	NA	% (se)	NA	% (se)	% (se)	NA
Additional "Pacific Islander" detailed responses	NA	% (se)	NA	% (se)	% (se)	NA
SOR Detailed Groups:						
Brazilian	% (se)	NA	NA	% (se)	% (se)	NA
Cape Verdean	% (se)	NA	NA	% (se)	% (se)	NA
Additional "SOR" detailed responses	NA	% (se)	NA	% (se)	% (se)	NA

Decision Criteria for Question 8: We will examine the results in Table 15 to evaluate the relative percentages of detailed reporting in each of the major categories:

- Detailed group reporting for detailed White groups and detailed Black groups should show similar levels across the treatments and/or higher levels in the combined approaches.
- Detailed reporting for detailed Hispanic groups, detailed Asian groups, and detailed NHPI) groups may be higher in the Separate approach for groups that have dedicated checkboxes (e.g., Mexican, Japanese, Samoan), and this is to be expected.
- Detailed reporting in the Separate approach for Hispanics, Asians, and NHPIs (formats which include dedicated detailed checkboxes for these groups) should be compared with the Combined approach with detailed checkboxes (web-based designs) and the multiple-detailed checkboxes approach for the Combined question (paper format Panel W).

9

Question 9. What is the effect of the different question formats on the reporting of detailed write-in responses in the appropriate write-in areas (e.g., reporting smaller groups that are not presented as checkboxes or examples)?

We will examine different question formats to determine how well "smaller" detailed groups are reported, in comparison to "larger" detailed groups which are employed as example/checkbox groups. For example, among Hispanic respondents, whether the percentage of smaller detailed groups (e.g., Bolivian) are collectively higher in formats where all groups utilize the same write-in area, compared with formats where only the "largest" detailed groups receive a detailed checkbox (e.g., Mexican, Puerto Rican, Cuban) and all other detailed groups must report via a write-in area. Similar comparisons will be conducted among detailed Asian responses and among detailed Pacific Islander responses.

The detailed groups shown in Figure 12 (below) are considered to be "larger" detailed groups for this analysis. This is because in some aspect of the question designs these groups are represented by dedicated detailed checkboxes, whereas other groups are not. Therefore, we want to explore whether there are more instances of these groups being reported when their respective dedicated checkbox is present, compared with when they, like other "smaller" detailed groups all have the option to report via a dedicated write-in area.

Figure 11. "Larger" Detailed Groups with Dedicated Checkboxes

German	African American	Lebanese
Irish	Jamaican	Iranian
English	Haitian	Egyptian
Italian	Nigerian	Syrian
Polish	Ethiopian	Moroccan
French	Somali	Algerian
Mexican or Mexican American	Chinese	Native Hawaiian
Puerto Rican	Filipino	Samoan
Cuban	Asian Indian	Chamorro
Salvadoran	Vietnamese	Tongan
Dominican	Korean	Fijian
Colombian	Japanese	Marshallese

Table 16. Reporting of "Larger" Detailed Groups and "Smaller" Detailed Groups by Treatment
(Percentage providing detailed responses)

	Separate Question		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
	Larger Detailed Groups	Smaller Detailed Groups	Larger Detailed Groups	Smaller Detailed Groups	Larger Detailed Groups	Smaller Detailed Groups
White	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic*	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian*	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
NHPI*	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

Decision Criteria for Question 9: Evaluate the reporting of detailed groups in each of the major categories to determine whether detailed write-in responses are being provided for not only larger groups which are presented as examples and/or detailed checkboxes, but also that smaller groups are being reported. Determine which question design approach yields more relevant data that encompasses detailed groups across myriad communities.

Question 10. How do the "race" reporting patterns for Hispanics compare across the different formats (Separate vs. Combined)?

It is important to understand race reporting patterns for Hispanic respondents across the different question formats. This will allow us to examine if there is a difference in Hispanics identifying in one of the different OMB race categories. For instance, in the combined question formats, is there a difference in Hispanics who also identify as Black ("Afro-Latinos") compared to the separate question format? These results will be shown in Table 17.

Table 17. Race Distribution for Hispanic Respondents by Treatment

	White alone	Black alone	Asian alone	AIAN alone	MENA alone	NHPI alone	SOR alone	Two or More	Invalid	Missing*
Separate Question	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Write-In Response Areas	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Combined Question with Detailed Checkboxes	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

* Missing includes responses of Hispanic alone, with no race response provided.

Decision Criteria for Question 10: We will evaluate results of the relative percentages for people in each of the major categories shown in the table above. We expect similar percentages of Black, Asian, AIAN, MENA, and NHPI groups, among Hispanic respondents, for the different formats.

3.9 | Testing a Middle Eastern or North African Category

Objective: To Evaluate the Use of a Distinct Middle Eastern and North African Category

Goal: Improve the collection and quality of data for Middle Eastern and North African populations, by introducing a distinct MENA category.

Another of the main objectives of the 2015 NCT is to evaluate the use of a distinct "Middle Eastern or North African" ("MENA") response category for collecting data on race/ethnicity. During its review of the original 1977 OMB Standards for Collecting Data on Race and Ethnicity in the mid-1990s, the OMB received a number of public comment recommendations to add a category for Arabs and Middle Easterners to the minimum groups listed in the standards. OMB did not accept this recommendation but encouraged further research on how to collect and improve data on this population group. The 2010 AQE was part of that research effort, conducting six focus groups with seventy-one participants of Middle Eastern and North African origin to understand more about their self-identity on census forms. The AQE focus groups sought to understand *how* and *why* people identify their race and ethnicity in different ways and in different contexts. The results from the focus groups indicated that many of the MENA participants had difficulty responding to the existing OMB race categories. They often did not know how to respond or felt left out. Also, the inclusion of the terms "Lebanese" and "Egyptian" as examples under the White racial category was viewed as wrong or incorrect by many of the AQE focus group respondents – both within the MENA focus groups, as well as across other focus groups. These comments often led to a recommendation by the focus group participants that there be a separate racial category for those who would identify as Middle Eastern, North African, or Arab.

In 2013, the Arab American Institute sent a letter to the Census Bureau and the OMB requesting a distinct category on the race or ethnic questions for people of Middle Eastern or North African origin. The letter was co-signed by 26 different organizations and scholars. In response to this request, the Census Bureau launched a comprehensive research and outreach program on the topic. As part of this, Census Bureau experts consulted with the OMB, key federal statistical agencies, professional demographic and sociological associations, academics, race and ethnicity experts, members of the Census Bureau's National Advisory Committee on Racial, Ethnic, and Other Populations (NAC), and MENA stakeholders on the classification and possible testing plans of a MENA category. In 2014, the NAC made a formal recommendation to the Census Bureau to test a distinct MENA category, and the Census Bureau decided to test this new MENA category in the 2015 NCT.

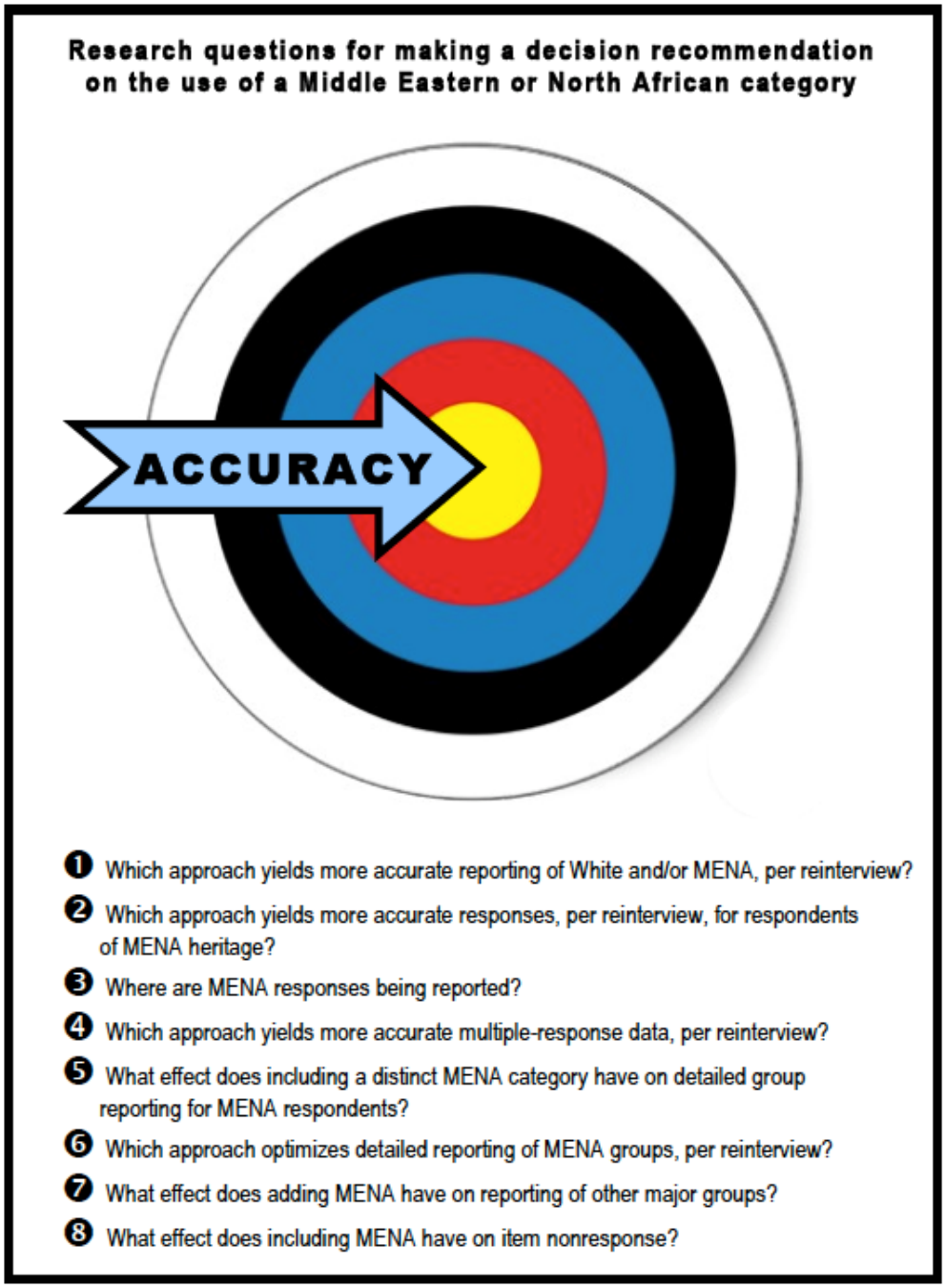
The 2015 NCT will evaluate the use of a "Middle Eastern or North African" (MENA) response category. In the survey, one-half of the 1.2 million households in the NCT will receive a question design that includes a distinct "Middle Eastern or North African" response category, accompanied by varying treatments of the other key dimensions being tested (separate questions approach vs. combined question approach; old instructions vs. new instructions; use of different terms – race, ethnicity, origin, or no terms at all). The other half of the 2015 NCT households will receive a question design where there is no distinct MENA category, and examples of MENA origin are listed among the examples for the White category.

The decision criteria for this dimension focus on evaluating the different approaches for including or not including a distinct MENA category. Specific analyses for this dimension will examine the research questions outlined over the next several pages, which will be used to tease out benefits and drawbacks of the various treatments. This will include examining the specific detailed responses for "Middle Eastern and North African" groups, including those within the MENA working classification list

(e.g., Lebanese, Egyptian, Iranian), as well as those groups for whom classification as MENA is unclear (e.g., Armenian, Turkish, Sudanese).

A recommendation on whether or not to include a distinct MENA category when collecting and producing data on race/ethnicity for the 2020 Census (MENA vs. NO MENA) will be made based on results from the 2015 NCT. An overview of the research questions that we are focusing on for this dimension are shown in the figure below.

Figure 12.



Within our analyses of the different research questions, the decisions regarding the dimension of a MENA response category predominantly rely on the following major factors: (a) better measures of "truth;" (b) lower item nonresponse to survey for MENA (reinterview) respondents; (c) levels of reporting detailed MENA groups; and (d) yielding lower usage of "Some Other Race" category. The foundation of determining which approach (MENA vs. NO MENA) is best, rests on measuring "truth." As described previously in Section 3.4, the reinterview will enable us to measure respondents' self-identified "true" racial and/or ethnic identity through a series of detailed questions and probes, which are compared to their responses on the self-response survey. Ultimately, these explorations will help us evaluate which approach yields more accurate and reliable results to reflect self-identification for respondents of Middle Eastern and North African heritage in the United States.

The overarching principle of this research is to find ways to improve the ***accuracy*** of data on race and ethnicity. Accuracy is the most important goal, and the research questions that we are examining yield insights to different aspects of accuracy. The eight research questions outlined in this section will explore different aspects of the ways in which we seek to improve the accuracy of data on race and ethnicity for respondents of MENA heritage in the United States. We note that all of the research questions are important, as they help us to understand the accuracy that we are measuring – whether it be the accuracy of reporting for major race/ethnic groups (such as MENA, White, Black, etc.), the accuracy of reporting multiple responses (such as White ***and*** MENA, Black ***and*** MENA, etc.), or the accuracy of reporting detailed nationalities or ethnic groups (such as Lebanese, Moroccan, etc.).

Pending the results of the 2015 National Content Test, the Census Bureau along with the Federal Interagency Working Group on Race and Ethnic Research will make recommendations to the Office of Management and Budget (OMB) on whether or not a MENA category should be included in the 2020 Census. Changes or modifications to the 1997 Race and Ethnicity Standards and relative decisions on how to classify and tabulate MENA responses will be decided upon by OMB. At this time, we will not speculate as to what those recommendations will be.

Strategies for Evaluating Success

Our strategies for evaluating the success of a distinct MENA category focus on the following series of key research questions. In our analyses, we will look to see which of the approaches work best (distinct MENA category vs. not including a MENA category). Within this, we focus on the following factors:

- Which approach yields the greatest accuracy?
- Which approach enables respondents to fully self-identify and reflect their "truth"?
- Which optimizes the reporting of multiple races and multiple ethnic origins?
- Which approach improves reporting of detailed nationalities and ethnic origins?
- Which approach reduces item nonresponse?
- Which approach improves reporting in major OMB categories (reducing "Some Other Race" reporting)?

The next several pages present a thorough, detailed walkthrough of the research questions, data tables, and decision criteria for analyzing the results for testing a distinct MENA category. Our hypotheses are presented as research questions (a priori), along with table shells and explanations of what will be analyzed. We also present decision criteria to explain how the results will be evaluated in order to make recommendations on the research question that is posed.

Question 1. Which approach (MENA vs. NO MENA) yields more accurate reporting as White and/or MENA, per the reinterview, for respondents of Middle Eastern or North African heritage?

The OMB definitions guide us in our classification, coding, and tabulation of detailed responses, and it is important to note that the definitions are geographically rooted. However, we must also recognize that the delineation and specificity of the OMB definitions for the geographic areas are not clearly defined. While "White" is defined as people with origins in the original peoples of Europe, the Middle East, and North Africa, the line for where these three areas begin and end is not clearly defined.

The Middle East, North Africa, and Europe all have countries with borders, which at times have shifted and merged, and some that are still contested. It is not definitively clear where "North Africa" ends and "Sub-Saharan Africa" begins, and where countries such as Sudan, Eritrea, Djibouti, Western Sahara, Somalia, should be classified. Similarly, it is unclear whether countries such as Turkey and Armenia, or countries that were part of the former Soviet Union are part of Europe, or part of the Middle East, or part of Asia. Additionally, the reference to "original peoples" in the OMB definitions add another layer of complexity for determining which groups are deemed the "original people" of these areas.

The Federal Register for the 2015 NCT received over 3,500 public comments expressing strong disagreement with the classification of Middle Eastern or North African as "White" – which is how these responses are currently classified in federal statistics on race. Addressing these questions is a challenge the Census Bureau cannot solve alone, and this will require work with OMB and external experts to understand the different perspectives and complexities. Insights from the 2015 NCT research will help us understand how respondents who report these groups self-identify their race/ethnicity.

We will examine the race/ethnic reporting of respondents of Middle Eastern and North African heritage to see which approach (MENA vs. NO MENA) yields more accurate reporting as White and/or MENA, per the reinterview. For responses provided in the reinterview we will compute the percentage that provided the same response category in the initial survey. The results of the 2015 NCT analyses, in conjunction with input from public comments and feedback from the 2015 MENA Expert Forum, as well as ongoing dialogues with OMB and external stakeholders and experts will guide MENA classification and tabulation decisions for the 2020 Census.

In the approaches with NO MENA category, we look to see whether MENA respondents either:

1. answer the race question by reporting a response within an OMB race category (i.e., White; Black; Asian; etc.),
2. leave the race question unanswered, or
3. provide a "MENA" response in the "Some Other Race" response area.

In the approach which includes a dedicated MENA category, we look to see whether MENA respondents:

1. also report a response within an OMB race category (i.e., White; Black; Asian; etc.),
2. only report a "MENA" response within the MENA response area, or
3. provide a "MENA" response in the "Some Other Race" response area.

Table 18. Reporting Patterns of the MENA Reinterview Population to the Survey Question Formats

	Question with Distinct MENA Category	Question with NO MENA Category
Identified as MENA ONLY	% (se)	% (se)
Identified as MENA *AND* White	% (se)	% (se)
Identified as MENA *AND* Black	% (se)	% (se)
Identified as MENA *AND* another group(s) (e.g., Asian, AIAN, etc.)	% (se)	% (se)
Identified as NOT MENA	% (se)	% (se)
Did not provide a response to survey	% (se)	% (se)
Reported an invalid response	% (se)	% (se)

Decision Criteria for Question 1: These response patterns are evaluated in conjunction with the NCT reinterview data to determine which approach (MENA or NO MENA) yields a more consistent overall race/ethnic distribution for respondents of Middle Eastern or North African heritage who report as White and/or MENA.

The inclusion of a distinct MENA category should yield more accurate reporting. Results from the reinterview will determine which approach (MENA category or NO MENA category), produces the most reliable and accurate responses for Middle Eastern and North African respondents, reflecting the respondent's "true" self-identified racial and ethnic identity. We will examine respondent answers to the "Yes" | "No" reinterview questions for White, MENA, and other categories in comparison with survey responses.

Question 2. Which approach (MENA vs. NO MENA) yields more accurate responses, per the reinterview, for respondents of Middle Eastern or North African heritage?

We will examine which approach (MENA vs. NO MENA) yields more accurate responses, per the reinterview, for respondents of Middle Eastern or North African heritage. For example, among respondents who self-identify as MENA in the reinterview, which approach obtains better matches with their survey response. The reinterview data will provide measures of response bias and gross difference rates, which we will examine in conjunction with respondent's "true" self-identified racial and ethnic identity as determined through the reinterview.

These data will enable us to evaluate the patterns of consistency for responses between the initial survey self-response and the reinterview responses for respondents of Middle Eastern or North African heritage. This will be a critical factor for determining which format (MENA vs. NO MENA) yields more accurate responses reflecting MENA respondent's "true" self-identified racial and ethnic identity.

Table 19. Percentage of Consistent Responses between Self-Response and Reinterview for Major Race/Ethnic Groups by Inclusion or Exclusion of Distinct MENA Category

	White	Hispanic	Black	Asian	AIAN	MENA	NHPI	SOR	Multiple
Question with Distinct MENA Category	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Question with NO MENA Category	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

Our analyses will also compare the levels of reporting for each race/ethnic response category in the survey and with those reported in the 2015 NCT reinterview. The reinterview data will provide measures of response bias (net difference rates), which we will examine in conjunction with respondent's "true" self-identified racial and ethnic identity as determined through the reinterview.

These data will enable us to evaluate the patterns of consistency for responses between the initial survey self-response and the reinterview responses for each of the major race/ethnic groups being studied. This will be a critical factor for determining which format (Separate or Combined) yields more accurate responses reflecting respondents' "true" self-identified racial and ethnic identity.

Table 20. Reinterview Race Distribution Compared to Self-Response

Question Design with Distinct MENA Category

	Reinterview								
Self-Response	**White**	**Hispanic**	**Black**	**Asian**	**AIAN**	**MENA**	**NHPI**	**SOR**	**Multiple**
White	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
NHPI	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Multiple	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Question Design with NO MENA Category

	Reinterview								
Self-Response	**White**	**Hispanic**	**Black**	**Asian**	**AIAN**	**MENA**	**NHPI**	**SOR**	**Multiple**
White	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
NHPI	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Multiple	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Decision Criteria for Question 2: Results from the reinterview will determine which question design approach (MENA vs. NO MENA) yields more accurate responses, per the reinterview, for respondents of Middle Eastern or North African heritage, producing the most reliable and accurate responses for reflecting respondents' "true" self-identified racial and ethnic identity. Among respondents who self-identify as MENA in the reinterview, we will evaluate which approach obtains better matches with their survey response.

The reinterview data will provide measures of response bias and gross difference rates, which we will examine in conjunction with respondent's "true" self-identified racial and ethnic identity as determined through the reinterview. These data will enable us to evaluate the patterns of consistency for responses between the initial survey self-response and the reinterview responses for respondents of Middle Eastern or North African heritage.

This will be a critical factor for determining which format (MENA vs. NO MENA) yields more accurate responses reflecting MENA respondent's "true" self-identified racial and ethnic identity. When a distinct MENA category is included, the item nonresponse rate for the race question should improve for respondents of Middle Eastern or North African heritage. We will evaluate this with reinterview data to examine patterns for self-identified MENA respondents.

❸ Question 3. Where are MENA responses being reported (for approaches with MENA vs. approaches with NO MENA category)?

This comparison regarding research question 3 focuses on the types of response patterns for respondents who report a MENA response, focusing on where these responses are entered. We examine their responses in the initial survey (prior to editing the data) to determine whether the inclusion or exclusion of a MENA category has an impact on where responses are reported.

Table 21. Reporting of MENA Responses by Treatment

		In which category was MENA response provided?				
	Total % Identified as MENA	**White**	**Black**	**MENA**	**SOR**	**Another category**
Distinct MENA category included (n = ?)	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
No MENA category (n = ?)	% (se)	% (se)	% (se)	N/A	% (se)	% (se)

Decision Criteria for Question 3: These response patterns are evaluated in conjunction with the NCT reinterview data to determine which approach (MENA or NO MENA) yields a more consistent overall race/ethnic distribution for respondents of Middle Eastern or North African heritage. In Table 21, we are looking for the approach with higher consistency for MENA respondents to determine which format enables Middle Eastern or North African respondents to fully report their MENA identity, reflecting lower “Some Other Race” reporting and more accurate responses, per the reinterview.

Question 4. Which approach (MENA vs. NO MENA) yields more accurate multiple-response data, per the reinterview (e.g., White *and* MENA; Black *and* MENA; Asian *and* MENA), for MENA respondents?

We will examine the overall level of multiple-responses between the survey and reinterview, for the approaches with and without a dedicated MENA category. We will examine whether MENA respondents also identify with another major response category (e.g., White, Black, Asian, etc.) or only report as MENA, and how those responses match with the extensive follow-up probing about each of the categories in the reinterview.

Table 22. Percentage of Consistent Multiple Responses between Self-Response and Reinterview for MENA Groups by Inclusion or Exclusion of Distinct MENA Category

	MENA alone	White *and* MENA	Black *and* MENA	Asian *and* MENA	Other multiple responses	NO Response
Question with Distinct MENA Category	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Question with NO MENA Category	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

Decision Criteria for Question 4: These response patterns are evaluated in conjunction with the NCT reinterview data to determine which approach (MENA or NO MENA) yields a more accurate multiple-response data, per the reinterview (e.g., White ***and*** MENA; Black ***and*** MENA; Asian ***and*** MENA), for respondents of Middle Eastern or North African heritage.

Question 5. What effect does including a distinct MENA category have on detailed group reporting for MENA respondents?

For the purpose of the 2015 NCT, the Census Bureau developed a working classification of the Middle Eastern or North African population in the United States. The working classification of MENA is based on the Census Bureau's on-going research and outreach efforts with community experts, stakeholders, and researchers. In addition, the Census Bureau has also documented how a wide range of organizations in the United States – including state and federal government agencies, research organizations, and universities – classify countries and territories from the Middle East or North Africa.

The Census Bureau's working MENA classification includes countries and territories that were in the majority of these MENA classifications. This approach classifies a person as MENA if they have ethnic origins or descent, roots, or heritage from any of the following nineteen countries: Algeria, Bahrain, Egypt, Iraq, Iran, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates, and Yemen. The following ethnicities are also included in the MENA classification: Amazigh or Berber, Arab, Assyrian, Bedouin, Chaldean, Copt, Druze, Kurdish, and Syriac.

Some of the experts in the Spring 2015 MENA Forum were concerned that countries such as Turkey, Sudan, or Somalia are not included in the current Census Bureau working classification of MENA. At the same time, however, other experts in the forum expressed concern that these groups would be included, and advised that they not be classified as MENA because they are not part of the Middle Eastern or North African geographic area. Obtaining this feedback was one of the main goals of the MENA Forum. We recognize that there are differing views on whether some countries are, or should be, part of the MENA category classification, and there are compelling justifications to both sides of this discussion. Therefore, for the purposes of the 2015 NCT research, we employ our current working MENA classification, and we will use this classification as the foundation for comparisons with other responses to the MENA category.

For analytical purposes, we will code all of the groups that people report to help us understand the types of responses that respondents provide with the new MENA category, as well as when the MENA category is not present. With all of the detailed disaggregated responses, we will obtain a more profound understanding of how various groups are reported and how they relate to the current MENA classification and examples.

Table 23. Reporting of Detailed MENA Groups in Different Category Response Areas for Treatments with Distinct MENA Category

Distinct MENA category included	In which category was the detailed MENA response provided?			
	White	**MENA**	Some Other Race	Another category
Lebanese	%	%	%	%
Iranian	%	%	%	%
Egyptian	%	%	%	%
Syrian	%	%	%	%
Moroccan	%	%	%	%
Algerian	%	%	%	%
[include other groups reported)	%	%	%	%
…	%	%	%	%

Table 24. Reporting of Detailed MENA Groups in Different Category Response Areas for Treatments with No MENA Category

NO MENA category	In which category was the detailed MENA response provided?			
	White	**MENA**	Some Other Race	Another category
Lebanese	%	N/A	%	%
Iranian	%	N/A	%	%
Egyptian	%	N/A	%	%
Syrian	%	N/A	%	%
Moroccan	%	N/A	%	%
Algerian	%	N/A	%	%
[include other groups reported)	%	N/A	%	%
…	%	N/A	%	%

Decision Criteria for Question 5: We will compare the reporting of MENA responses, in terms of which category the detailed MENA responses was provided, to examine whether major differences of reporting occur. Specifically, we are interested in whether and where detailed MENA responses are provided, when no distinct MENA category is presented.

For example, we will examine whether responses such as Lebanese, Egyptian, or Iranian, are reported in connection with the Some Other Race response category when no distinct MENA category is presented, and/or if MENA responses are reported in other category areas such as White, Black, Asian, etc. Conversely, we will examine where these types of MENA responses are reported when there **is** a dedicated MENA category present.

Additionally, we will examine reporting patterns for groups that are not currently in our MENA classification (e.g., Afghans, Armenians, Sudanese, Turkish, etc.) but may have some respondents who consider themselves to be MENA and would report as such in their response.

Overall, we expect that the inclusion of a MENA category and detailed examples of Middle Eastern and North African heritage should improve detailed reporting for MENA respondents. We will look to see which approach yields higher levels of detailed MENA reporting (the approach with no distinct MENA category vs. the approach with a distinct MENA category).

6 **Question 6. Which approach (MENA vs. NO MENA) best optimizes detailed reporting of Middle Eastern or North African groups, per the reinterview?**

We will examine the different approaches (MENA vs. NO MENA) to determine which approach best optimizes detailed reporting of Middle Eastern or North African groups, per the reinterview. Examining the two formats, we will determine how the larger Middle Eastern groups in the United States (e.g., Lebanese, Iranian, Egyptian, etc.) and the larger North African groups in the United States (e.g., Syrian, Moroccan, Algerian, etc.) are reported. In addition, we will examine how smaller detailed Middle Eastern groups and North African groups are reported, in comparison to the larger detailed Middle Eastern groups and North African groups which are employed as example/checkbox groups in the question designs.

Table 25. Reporting of Detailed MENA Groups by Treatment

	Dedicated MENA Category		No MENA Category	
	Percent in Survey	Number in Reinterview	Percent in Survey	Number in Reinterview
Lebanese	% (se)	#	% (se)	#
Iranian	% (se)	#	% (se)	#
Egyptian	% (se)	#	% (se)	#
Syrian	% (se)	#	% (se)	#
Moroccan	% (se)	#	% (se)	#
Algerian	% (se)	#	% (se)	#
[insert next largest groups]	% (se)	#	% (se)	#
…	% (se)	#	% (se)	#
…	% (se)	#	% (se)	#
…	% (se)	#	% (se)	#
…	% (se)	#	% (se)	#

Decision Criteria for Question 6: The question approach that yields more accurate detailed data, per the reinterview will help inform which design format to utilize. In the end, our 2015 NCT research will provide critical information for illuminating how respondents with origins from various parts of the Middle East, North Africa, and adjacent areas, respond to different versions of the questions that include, or do not include, a distinct MENA category. Coupled with insights from the NCT reinterview, we will ascertain how and where respondents self-identify and how this compares to the feedback we received from the 2015 MENA Forum and from the thousands of Federal Register public comments on the classification of the MENA category.

These results and our ongoing outreach and dialogue with stakeholders will help inform recommendations on the inclusion of a MENA category and the classification of responses. The inclusion of a MENA category and detailed examples of Middle Eastern and North African heritage should improve detailed reporting for MENA respondents. The level of detailed MENA reporting should be higher than the approach without a distinct MENA category.

7 **Question 7. What effect does adding a distinct MENA category have on the reporting of other major racial/ethnic groups (e.g., White, Hispanic, Black, Asian, etc.) and the reporting of "Some other race or ethnicity?"**

We will examine the two treatments for testing this dimension: (1) designs that include a dedicated MENA category; and (2) designs with no separate MENA category. For designs where a MENA category is placed within the current category lineup, the "White" example groups are revised and the Middle Eastern and North African examples of Lebanese and Egyptian are replaced with the European examples of Polish and French. The MENA checkbox category will have the examples of Lebanese, Iranian, Egyptian, Syrian, Moroccan, and Algerian. All other checkbox categories and write-in spaces remain the same.

We will compare the reporting of MENA responses, as well as the relative levels of reporting for other major categories, primarily focusing on the reporting of "White," "Black," or "Asian" to examine whether major differences of reporting occur. Specifically, we are interested in whether detailed MENA responses are provided in response to other race/ethnic category areas, when no distinct MENA category is presented. For example, we will examine whether responses such as Lebanese, Egyptian, or Iranian, are reported in connection with the Some Other Race response category when no distinct MENA category is presented, and/or if MENA responses are reported in other category areas such as White, Black, or Asian. Conversely, we will examine where these types of MENA responses are reported when there **is** a dedicated MENA category present.

We will examine the estimated percentages of people in each of the following major categories:

- White
- Hispanic or Latino (Hispanic)
- Black or African American (Black)
- Asian
- American Indian or Alaska Native (AIAN)
- Middle Eastern or North African (MENA)
- Native Hawaiian or Other Pacific Islander (NHPI)
- Some Other Race, Ethnicity, or Origin (SOR)
- Multiple Responses

Table 26. Race/Ethnic Group Distribution by Treatment

	Distinct MENA Category Included	No MENA Category
White	% (se)	% (se)
Hispanic	% (se)	% (se)
Black	% (se)	% (se)
Asian	% (se)	% (se)
AIAN	% (se)	% (se)
MENA	**% (se)**	**% (se)**
NHPI	% (se)	% (se)
SOR	% (se)	% (se)
Multiple	% (se)	% (se)

Decision Criteria for Question 7: The distributions of race and ethnic groups will be examined for both approaches. We will compare the reporting of MENA responses, as well as the relative levels of reporting for other major categories, primarily focusing on the reporting of "White," "Black," or "Asian" to examine whether major differences of reporting occur. The percentage of respondents of Middle Eastern or North African heritage should be higher when a distinct MENA category is presented as one of the response options. We will evaluate impact of adding a MENA category to the relative percentages of people in each of the major race/ethnic categories:

- The percentage of people reporting MENA responses in the SOR category should be lower when a distinct MENA category is included.
- The percentage of MENA responses in the White category should be similar or lower when a MENA category is included, compared with when MENA examples are listed as part of the White category.
- Similar percentages of Hispanic, Black, Asian, AIAN, and NHPI groups should be seen, regardless of the different format.

Question 8. What effect does including a MENA category have on item nonresponse?

We will examine whether the item nonresponse to the race question is higher or lower on panels where a dedicated MENA category is included. Two potential examinations are possible here. For the first, we will examine the overall level of item nonresponse between the two formats. For the second, we will delve further into the analyses with the insights from the reinterview data.

Table 27. Pattern of Response by Presence of Distinct MENA Category

	Item Nonresponse Rate
Distinct MENA Category	% (se)
No MENA Category	% (se)

Decision Criteria for Question 8: Nonresponse rates will be examined for both approaches (MENA vs. NO MENA). The overall item nonresponse rate should remain stable, or improve, with the addition of a MENA category.

3.10 | Testing Alternative Instructions and Terminology

Objective: To Improve the Wording of the Instructions and Improve Question Terminology

Goal: Improve the understanding of the question and optimize the reporting of multiple-responses for respondents of multiracial or multiethnic heritage, by introducing new instructions and alternative terminology for the race/ethnicity questions.

Another objective of the 2015 NCT is to evaluate the use of new instruction wording and alternative terminology for the question format approaches for collecting data on race and ethnicity. This research is being undertaken to improve the clarity of the question and make it more apparent that more than one group may be selected and to enable respondents, especially multiracial and multiethnic respondents, to more easily self-identify in ways that reflect how they see themselves. This objective builds upon the successful findings of the 2010 AQE research which showed promising strategies for allowing respondents to report all of the groups with which they self-identify. The AQE research found that combining race and ethnicity into one question provided a cleaner conceptual format and respondents provide higher levels of multiple-group reporting, which were confirmed in the reinterview, and more accurately reflected respondent's self-identification.

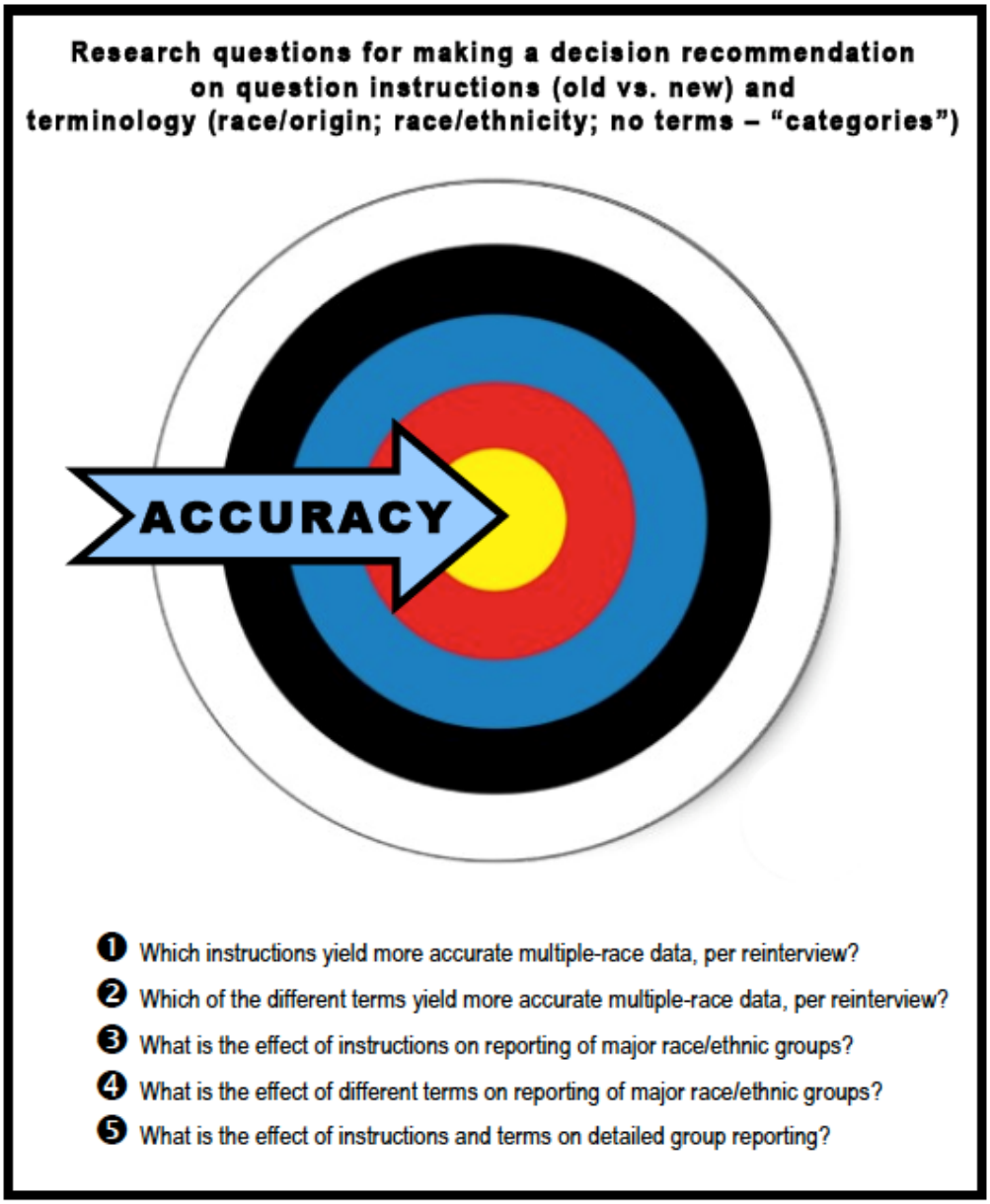
In the 2015 NCT, the different instructions and terminology are being tested in various ways. Both of these approaches will be tested in the 2015 NCT, with new data collection methods, including Internet, smartphone, and responses with telephone questionnaire assistance.

First, the research will evaluate the use of different approaches for the instruction wording used to collect data on race and ethnicity. The 2010 AQE research found that respondents frequently overlook the instruction to "*Mark* [***X***] *one or more boxes*" and have difficulty understanding the instructions. From the 2010 AQE qualitative research we learned that some respondents stop reading the instruction after noticing the visual cue [***X***] and proceed directly to do just that – mark a box – overlooking the remainder of the instruction. The new instruction being tested in the 2015 NCT ("*Mark all boxes that apply*") is an attempt to improve the clarity of the question and make it more apparent that more than one group may be selected ("*Note, more than one group may be selected*").

Second, the 2015 NCT research will evaluate the use of different conceptual terms (e.g., race, origin, ethnicity, or no terms) in the wording of questions for collecting data on race and ethnicity. The use of "race" and "origin" as terminology (old instructions) will be used to guide respondents to answer the question (e.g., "*What is Person 1's race or origin?*"). One alternative option being explored tests the use of both the terms "ethnicity" along with "race" in the question stem and/or instructions (e.g., "*What is Person 1's race or ethnicity?*"). A second alternative option being explored tests the removal of the terms "race," "origin," and "ethnicity" from the question stem and instructions. Instead, a general approach asks, "Which categories describe Person 1?" These options are being tested to determine whether we can improve the understanding of the question concept and reduce confusion among respondents by using different terms (or no terms at all) for the race and ethnicity questions.

The decision criteria for the Instruction Wording Dimension and for the Alternative Terminology Dimension focus on evaluating the different approaches for instructions and terminology. Specific analyses for these dimensions will examine the research questions outlined in this section, which will be used to tease out benefits and drawbacks of the various treatments. A recommendation on the best instructions and the best terminology for collecting and producing data on race/ethnicity for the 2020 Census will be made based on results from the 2015 NCT. An overview of the research questions that we are focusing on for these dimensions are shown in the figure below.

Figure 13.



The overarching principle of this research is to find ways to improve the ***accuracy*** of data on race and ethnicity. Accuracy is the most important goal, and the research questions that we are examining yield insights to different aspects of accuracy. The five research questions outlined in this section will explore different aspects of the ways in which we seek to improve the accuracy of data on race and ethnicity. We note that all of the research questions are important, as they help us to understand the accuracy that we are measuring – whether it be the accuracy of reporting for major race/ethnic groups (such as White, Hispanic, Black, Asian, etc.), the accuracy of reporting multiple responses (such as White ***and*** Black, Black ***and*** American Indian, White ***and*** Asian ***and*** Pacific Islander, etc.), or the accuracy of reporting detailed nationalities or ethnic groups (such as Irish, Puerto Rican, Jamaican, Filipino, etc.).

Within our analyses of the different research questions, the decisions regarding these dimensions of instructions and alternative terminology predominantly rely on the following major criteria: (a) better measures of "truth" for respondents of multiple races or multiple ethnic origins; (b) yielding self-identified reporting within OMB categories; and (c) optimizes detailed reporting.

As described previously in Section 3.4, the reinterview will enable us to measure respondents' self-identified "true" racial and/or ethnic identity through a series of detailed questions and probes, which are compared to their responses on the self-response survey. Ultimately, these explorations will help us evaluate which question format yields more accurate and reliable results to reflect respondents' self-identification.

Strategies for Evaluating Success

Our strategies for evaluating the success of a different instruction wording and alternative terminology focus on the following series of key research questions. In our analyses, we will look to see which of the approaches work best (OLD instructions vs. NEW instructions) (race/origin vs. race/ethnicity vs. no terms – "categories). Within this, we focus on the following factors:

- Which approach yields the greatest accuracy for multiple responses?
- Which approach enables respondents to fully self-identify and reflect their "truth"?
- Which optimizes the reporting of multiple races and multiple ethnic origins?
- Which approach improves reporting of detailed nationalities and ethnic origins?
- Which approach improves reporting in major OMB categories (reducing "Some Other Race" reporting)?

The next several pages present a thorough, detailed walkthrough of the research questions, data tables, and decision criteria for analyzing the results for instructions and alternative terminology. Our hypotheses are presented as research questions (a priori), along with table shells and explanations of what will be analyzed. We also present decision criteria to explain how the results will be evaluated in order to make recommendations on the research question that is posed.

Question 1. Which instructions (old vs. new) yield more accurate multiple-race data, per the reinterview (e.g., White *and* Black; White *and* Asian; etc.)?

We will examine the overall level of accuracy for multiple-responses (as a total) between the survey and the reinterview. This will be done by comparing the question approach with old instructions ("*Mark* [***X***] *one or more boxes*") vs. the question approach with new instructions ("*Mark all boxes that apply*") to examine overall consistency.

Table 28. Overall Consistency of Self-Response and Reinterview for Multiple-Responses by Question Format

	Consistency of Multiple-Responses (Survey vs. Reinterview)
OLD: "*Mark [X] one or more boxes*"	% (se)
NEW: "*Mark all boxes that apply*"	% (se)

In addition, we will examine the major multiple-response combination groups from the 2015 NCT reinterview in comparison to what was reported in the self-response survey. We may also examine other multiple-response combinations greater than 1 percent. For responses provided in the reinterview we will compute the percentage that provided the same response category in the initial survey. Based on results from 2010 Census and 2010 AQE, the expected major multiple-response categories for this analysis are:

1. White ***and*** Black
2. White ***and*** Hispanic
3. White ***and*** Asian
4. White ***and*** AIAN
5. Hispanic ***and*** Black

Table 29. Consistency of Self-Response and Reinterview for Major Multiple-Responses Groups by Question Format

Consistency of Multiple-Responses (Survey vs. Reinterview)			
OLD: "*Mark [X] one or more boxes*"		**NEW:** "*Mark all boxes that apply*"	
White ***and*** Black	% (se)	White ***and*** Black	% (se)
White ***and*** Hispanic	% (se)	White ***and*** Hispanic	% (se)
White ***and*** Asian	% (se)	White ***and*** Asian	% (se)
White ***and*** AIAN	% (se)	White ***and*** AIAN	% (se)
Hispanic ***and*** Black	% (se)	Hispanic ***and*** Black	% (se)

Table 30. Reinterview Race Reporting Compared to Self-Response for Major Multiple Response Groups

OLD: "*Mark [X] one or more boxes*"

				Reinterview		
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Single Response
White ***and*** Black	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White ***and*** Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White ***and*** Asian	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
White ***and*** AIAN	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Hispanic ***and*** Black	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single Response	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

NEW: "*Mark all boxes that apply*"

				Reinterview		
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Single Response
White ***and*** Black	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White ***and*** Hispanic	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White ***and*** Asi**an**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
White ***and*** AIAN	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Hispanic ***and*** Bla**ck**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single Response	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Decision Criteria for Question 1: Results from the reinterview will determine which instructions (Old vs. New), produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial and ethnic identity. We will evaluate results to determine which format enables multiracial respondents to fully report their multiple identities, reflecting lower "Some Other Race" reporting and more accurate responses, per the reinterview.

The percentage of multiple-responses should be similar for the different formats, or greater for the "Mark all that apply" instructions. If multiple-response reporting is higher with this approach for groups such as White ***and*** Black, White ***and*** AIAN, White ***and*** Asian, etc. this would indicate a favorable design.

Question 2. Which of the different terms yield more accurate multiple-response data (e.g., White *and* Black; White *and* Asian; etc.), per the reinterview?

Results from the reinterview will determine which terms (race/ethnicity, race/origin, or no terms at all), produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial and ethnic identity.

We will examine the overall level of accuracy for multiple-responses (as a total) between the survey and the reinterview. This will be done by comparing the question approaches which use different terms to examine overall consistency: race/origin; race/ethnicity; and no terms at all ("categories").

Table 31. Overall Consistency of Self-Response and Reinterview for Multiple-Responses by Question Format

Terms	Consistency of Multiple-Responses (Survey vs. Reinterview)
race/origin	% (se)
race/ethnicity	% (se)
no terms at all ("categories")	% (se)

In addition, we will examine the major multiple-response combination groups and compare their percentages to the 2015 NCT reinterview (we may also examine other multiple-response combinations greater than 1 percent). Based on results from 2010 Census and 2010 AQE, the expected major multiple-response categories for this analysis are:

1. White ***and*** Black
2. White ***and*** Hispanic
3. White ***and*** Asian
4. White ***and*** AIAN
5. Hispanic ***and*** Black

Table 32. Consistency of Self-Response and Reinterview for Major Multiple-Responses Groups by Question Format

	Race/ Origin	Race/ Ethnicity	No terms at all ("categories")
White ***and*** Black	% (se)	% (se)	% (se)
White ***and*** Hispanic	% (se)	% (se)	% (se)
White ***and*** Asian	% (se)	% (se)	% (se)
White ***and*** AIAN	% (se)	% (se)	% (se)
Hispanic ***and*** Black	% (se)	% (se)	% (se)

Table 33. Self-Response Race Distribution Compared to Reinterview Race Distribution

"Race/Origin"

	Reinterview					
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Single Response
White *and* **Black**	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Hispanic**	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White *and* **Asian**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
White *and* **AIAN**	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Hispanic *and* **Black**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single Response	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

"Race/Ethnicity"

	Reinterview					
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Single Response
White *and* **Black**	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Hispanic**	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White *and* **Asian**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
White *and* **AIAN**	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Hispanic *and* **Black**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single Response	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

"No Terms at All" ("Categories")

					Reinterview	
Self-Response	White ***and*** Black	White ***and*** Hispanic	White ***and*** Asian	White ***and*** AIAN	Hispanic ***and*** Black	Single Response
White *and* **Black**	**% (se)**	% (se)	% (se)	% (se)	% (se)	% (se)
White *and* **Hispanic**	% (se)	**% (se)**	% (se)	% (se)	% (se)	% (se)
White *and* **Asian**	% (se)	% (se)	**% (se)**	% (se)	% (se)	% (se)
White *and* **AIAN**	% (se)	% (se)	% (se)	**% (se)**	% (se)	% (se)
Hispanic *and* **Black**	% (se)	% (se)	% (se)	% (se)	**% (se)**	% (se)
Single Response	% (se)	% (se)	% (se)	% (se)	% (se)	**% (se)**

Decision Criteria for Question 2: Results from the reinterview will determine which terms (race/origin, race/ ethnicity, or no terms at all), produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial/ethnic identity.

We will evaluate results to determine which format enables multiracial respondents them to fully report their multiple identities, reflecting lower "Some Other Race" reporting and more accurate responses, per the reinterview.

If multiple-response reporting is higher for groups such as White ***and*** Black, White ***and*** AIAN, White ***and*** Asian, etc. this would indicate a favorable design.

❸ Question 3. What is the effect of the instructions (old vs. new) on the reporting of major racial and ethnic groups (e.g., White, Hispanic, Black, Asian, etc.), for yielding more accurate responses, per the reinterview?

Results from the reinterview will determine which instructions (old vs. new), produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial and ethnic identity.

We will examine the overall level of accuracy for the major race/ethnic categories between the survey and the reinterview. This will be done by comparing the question approaches which use different terms to examine overall consistency.

Table 34. Consistency of Self-Response and Reinterview for Major Response Groups by OLD Instructions vs. NEW Instructions Approach

Consistency of Major Race/Ethnic Groups (Survey vs. Reinterview)			
OLD: "*Mark [X] one or more boxes*"		**NEW:** "*Mark all boxes that apply*"	
White	% (se)	White	% (se)
Hispanic	% (se)	Hispanic	% (se)
Black	% (se)	Black	% (se)
Asian	% (se)	Asian	% (se)
AIAN	% (se)	AIAN	% (se)
MENA	% (se)	MENA	% (se)
NHPI	% (se)	NHPI	% (se)
SOR	% (se)	SOR	% (se)
Multiple Responses	% (se)	Multiple Responses	% (se)

Decision Criteria for Question 3: Evaluate the effect of the different instructions on the reporting of major racial and ethnic groups (e.g., White, Hispanic, Black, Asian, etc.), per the reinterview.
Results from the reinterview will determine which instructions (old vs. new) produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial and ethnic identity.

4 **Question 4. What is the effect of the different terms on the reporting of major racial and ethnic groups (e.g., White, Hispanic, Black, Asian, etc.), for yielding more accurate responses, per the reinterview?**

Results from the reinterview will determine which terms (race/origin, race/ethnicity, or no terms at all), produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial and ethnic identity.

We will examine the overall level of accuracy for the major race/ethnic categories between the survey and the reinterview. This will be done by comparing the question approaches which use different terms to examine overall consistency.

Table 35. Consistency of Self-Response and Reinterview for Major Response Groups by Question Terms

Consistency of Reporting (Survey vs. Reinterview)					
Race/Origin		**Race/Ethnicity**		**No terms at all ("categories")**	
White	% (se)	White	% (se)	White	% (se)
Hispanic	% (se)	Hispanic	% (se)	Hispanic	% (se)
Black	% (se)	Black	% (se)	Black	% (se)
Asian	% (se)	Asian	% (se)	Asian	% (se)
AIAN	% (se)	AIAN	% (se)	AIAN	% (se)
MENA	% (se)	Asian	% (se)	MENA	% (se)
NHPI	% (se)	AIAN	% (se)	NHPI	% (se)
SOR	% (se)	SOR	% (se)	SOR	% (se)
Multiple Responses	% (se)	Multiple Responses	% (se)	Multiple Responses	% (se)

Decision Criteria for Question 4: Evaluate the effect of the different terms on the reporting of major racial and ethnic groups (e.g., White, Hispanic, Black, Asian, etc.), per the reinterview. Results from the reinterview will determine which terms (race/origin, race/ethnicity, or no terms at all), produce the most reliable and accurate responses for reflecting the respondent's "true" self-identified racial and ethnic identity.

5 **Question 5. What is the effect of the instructions and terms on detailed group reporting (e.g., detailed reporting for Whites; detailed reporting for Hispanics; detailed reporting for Blacks; detailed reporting for Asians; etc.), in terms of optimizing detailed reporting?**

Results from the survey response will determine which combination of terms (race/origin, race/ethnicity, or no terms at all) and instructions (old vs. new), produce the most detailed reporting for each major race and ethnic group.

Table 36. Detailed Reporting for Major Race/Ethnic Groups by Terms and Instructions
(Percentage providing detailed responses)

DETAILED REPORTING	RACE/ORIGIN TERMS		RACE/ETHNICITY TERMS		NO TERMS AT ALL ("Categories")	
	OLD Instructions	NEW Instructions	OLD Instructions	NEW Instructions	OLD Instructions	NEW Instructions
White	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Hispanic*	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Black	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
Asian*	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
AIAN	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
MENA	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
NHPI*	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)
SOR	% (se)	% (se)	% (se)	% (se)	% (se)	% (se)

Decision Criteria for Question 5: Examine the effect of the instructions and terms for maximizing detailed reporting in each major group.

3.11 | Testing Performance of Questions in Paper and Web-Based Designs

Objective: To Evaluate Performance of Questions for Paper and in Web-Based Designs

Goal: Improve race/ethnic reporting through use of enhanced question designs.

The 2015 NCT provides the critical opportunity to compare the success of different question designs to determine how they perform in new web-based data collection methods using the Internet, smartphone, and telephone response options. With the advantage of new technology to collect data via web-based designs, we are testing different versions of the Internet questions with multiple screens for collecting responses, and the inclusion of dedicated write-in areas and/or detailed checkboxes for soliciting detailed racial and ethnic origins.

Different designs for the paper formats are also being tested in the 2015 NCT to see how they operate. However, following the goals of a re-engineered 2020 Census, our main focus is on testing the fully factorial components of each dimension via web-based designs. Each component is included in the various paths of the web-based designs so that every scenario is tested for:

- separate vs. combined with write-ins vs. combined with detailed checkboxes
- distinct MENA category vs. NO MENA category
- old instructions vs. new instructions
- race/origin vs. race/ethnicity vs. no terms at all – “categories”

This overarching objective of the 2015 NCT research on race and ethnicity will enable us to evaluate the use of enhanced question designs for both paper-based approaches and web-based approaches to collect data on race and ethnicity. We will examine the results described in detail in the previous sections to determine which design versions perform better than others.

Overall, this research is being undertaken to improve the clarity of the question(s) and to enable respondents to report all of the groups with which they self-identify. As the decisions for the different research dimensions are made, they will guide us to a pointed outcome on which question design performs best.

To illustrate this, we present the following graphical series to explain how a question design decision unfolds in accordance with the test results. To begin, we start with the recognition that there are 36 different paths for testing each of the different research dimension configurations. Each one of the 36 paths are illustrated by the boxes below, which are numbered 1 through 36:

1	2	13	14	25	26
3	4	15	16	27	28
5	6	17	18	29	30
7	8	19	20	31	32
9	10	21	22	33	34
11	12	23	24	35	36

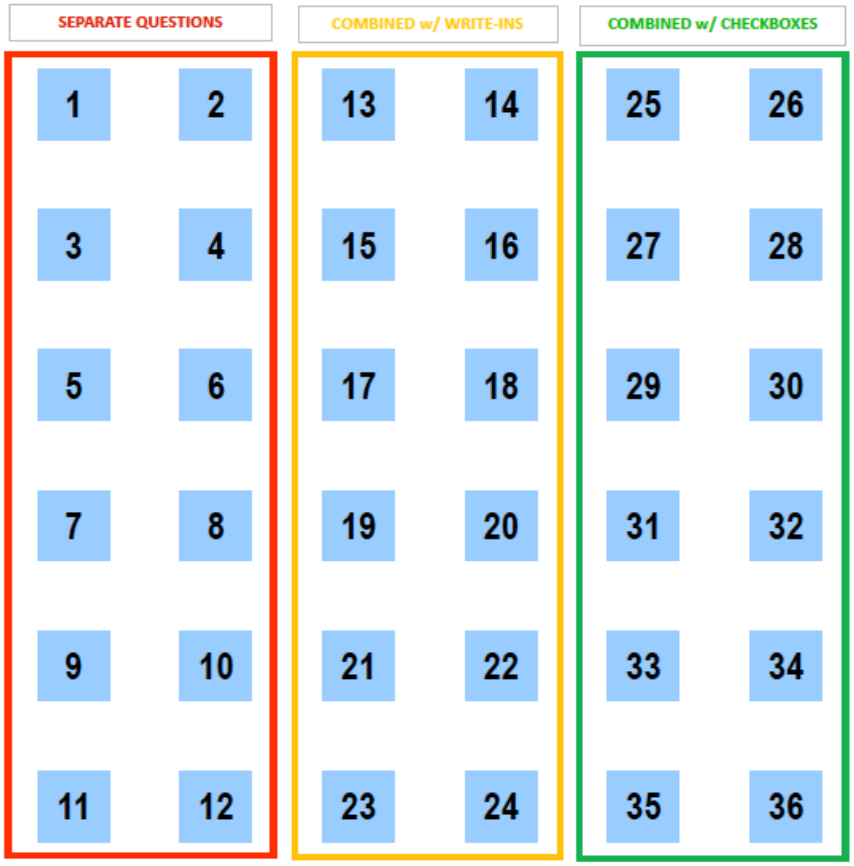
Testing Options for Separate vs. Combined with Write-Ins vs. Combined with Detailed Checkboxes.

First, the research will evaluate the analytical questions for question format dimension (separate questions vs. combined question with write-in areas vs. combined question with detailed checkboxes). This will determine which question format performs best, according to the research questions and decision criteria (which are detailed in section 3.7.2). Highlighted below by the different colored lines are the three pieces of this research dimension.

The first third of the 36 options (the options numbered 1-12) employ a design that tests the separate questions approach (these options are outlined in red).

The next third of the 36 options (the options numbered 13-24) employ a design that tests the combined question approach with write-in areas for collecting detailed responses (these options are outlined in yellow).

The last third of the 36 options (the options numbered 25-36) employ a design that tests the combined question approach with detailed checkboxes for collecting detailed responses (these options are outlined in green).

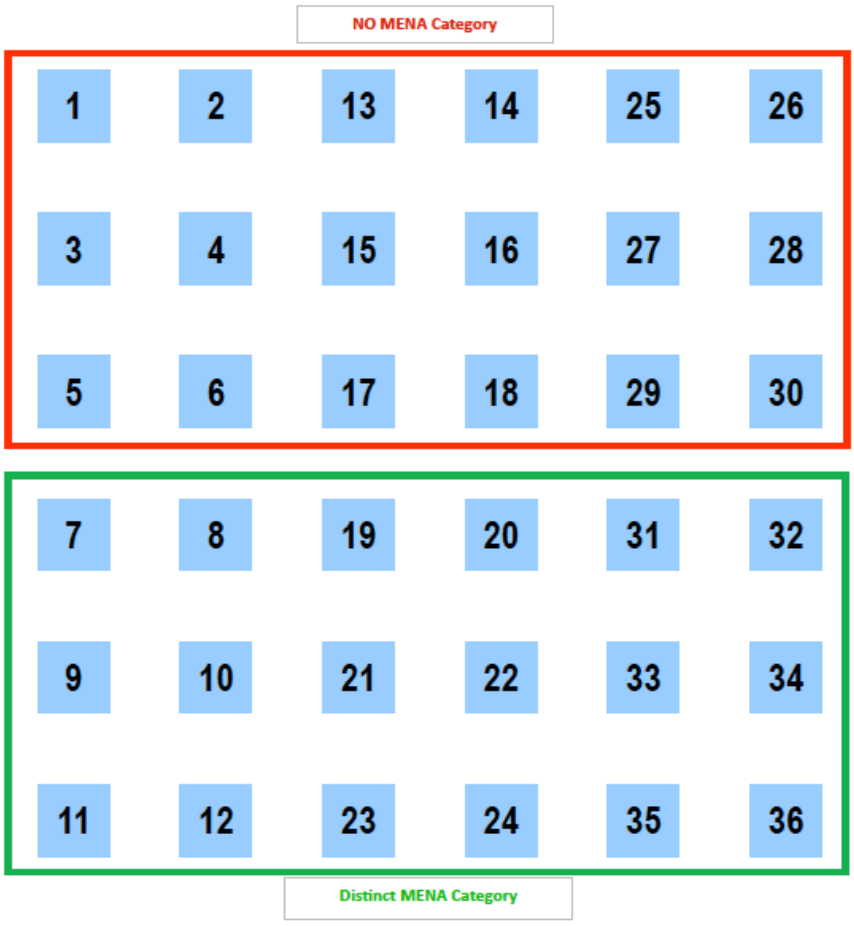


Based on the research findings, a recommendation will be made for employing one of these three options for question format (separate vs. combined with write-ins vs. combined with detailed checkboxes).

Testing Options for NO MENA Category vs. MENA Category.

Next, the research will evaluate the questions for the dimension of testing a MENA category (No MENA category vs. MENA category). Our analyses will examine which category structure performs best, according to the research questions and decision criteria (which are detailed in section 3.7.3). The two components of this dimension are outlined below by the red line and the green line below. Overall, one-half of the households in the NCT were presented with a question design that included a distinct "Middle Eastern or North African" response category, accompanied by varying treatments of the other key dimensions being tested (separate questions approach vs. combined question approach; old instructions vs. new instructions; and use of different terms – race, ethnicity, origin, or no terms at all).

The top half of the 36 options in the graphic (the options numbered 1-6, 13-18, and 25-30) all test the design approach without a distinct MENA category (these options are outlined in red).



The bottom half of the 36 options (the options numbered 7-12, 19-24, and 31-36) all test the design approach where a distinct MENA category is included in the question design (these options are outlined in green). This depicts the other half of the 2015 NCT households, which received a question design where there was no distinct MENA category. Instead, the examples of MENA origins were listed among the examples for the White category.

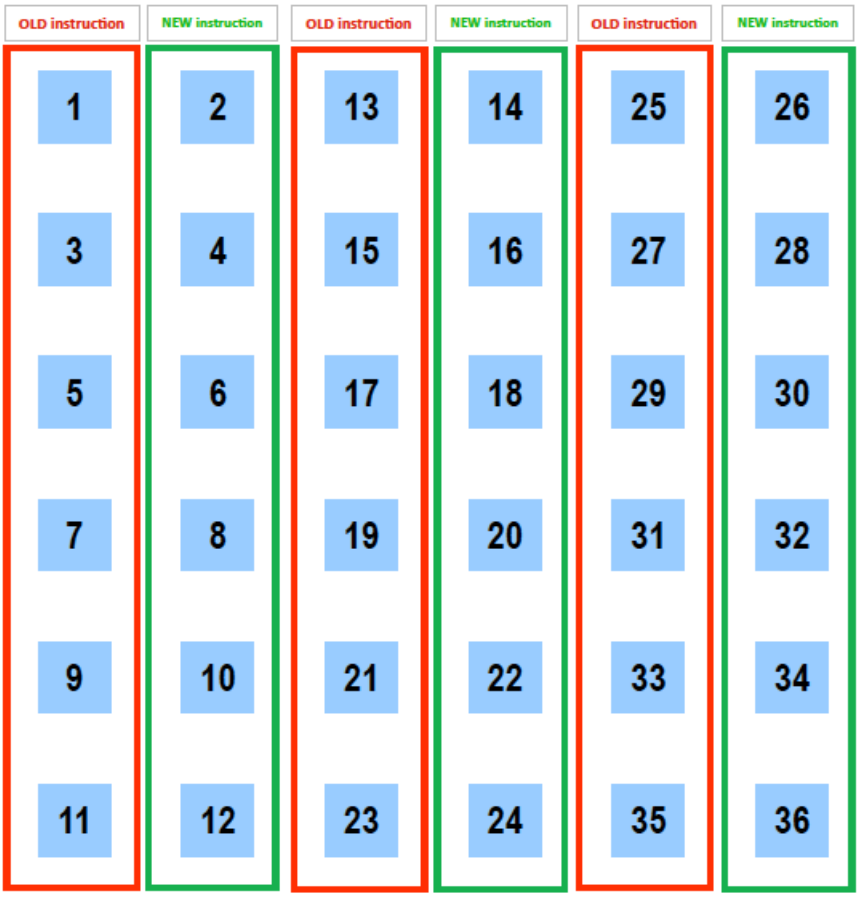
Based on the research findings, a recommendation will be made for employing one of these two options for respondent categories – no distinct MENA category vs. utilizing a distinct MENA category.

Testing Options for Old Instructions vs. New Instructions.

Additionally, the research will evaluate the questions for the dimension regarding the use of different approaches for the question instruction wording used to collect data on race and ethnicity. With respect to instruction wording (old vs. new), the different instructions are analytically examined to determine which format performs best, according to the research questions and decision criteria presented in this study plan (which are detailed in section 3.7.4).

In the graphic below, the red lines and green lines outline the pieces of this dimension. Alternating through the sequence of question designs, half of the designs (the options numbered 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, and 35) employ the old instructions to "*Mark* [*X*] *one or more boxes*" (these options are outlined in red).

The other half of the designs (the options numbered 2, 4, 6,8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, and 36) employ the new instructions to "*Mark all boxes that apply*" and "*Note, more than one group may be selected*" (these options are outlined in green).

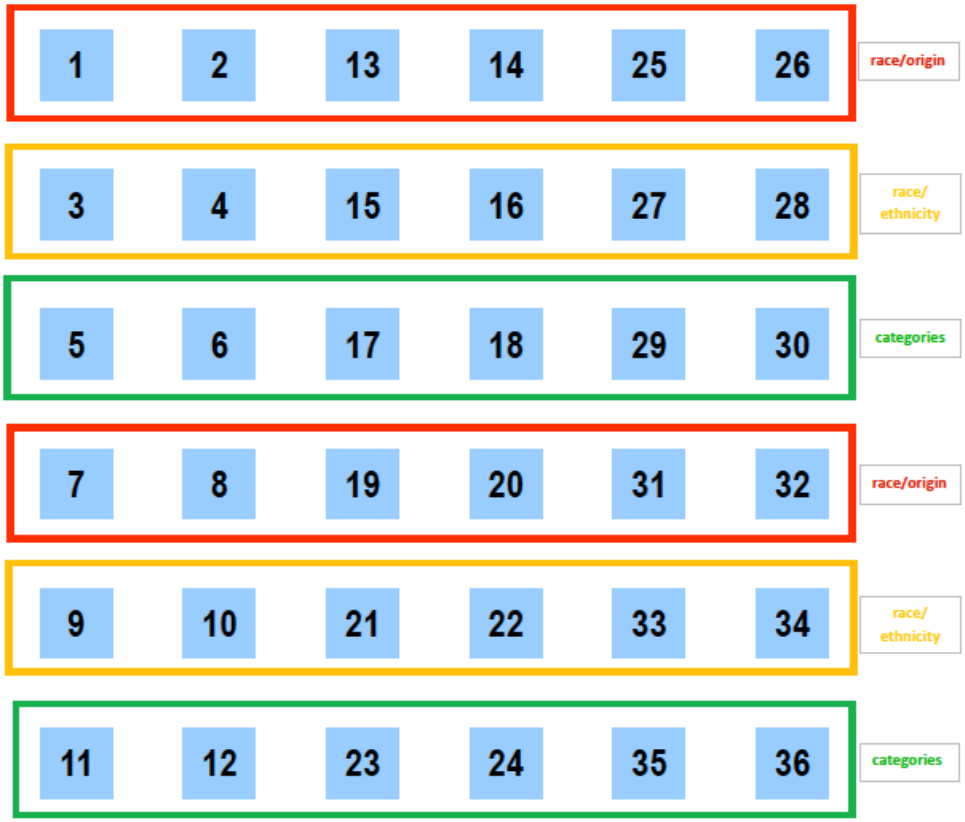


Based on the research findings, a recommendation will be made for employing one of the two options for respondent categories (old instructions vs. new instructions).

Testing Options for Race/Origin Terms vs. Race/Ethnicity Terms vs. No Terms ("Categories").

The research also evaluates the analytical questions for the dimension regarding the use of different terminology to collect data on race and ethnicity. We examine which format performs best (race/origin vs. race/ethnicity vs. no terms at all – "categories") based on the research questions and decision criteria outlined in this study plan (which are detailed in section 3.7.4). The three pieces of this research dimension are highlighted in the graphic below by the red, yellow, and green lines.

Alternating through the design sequences, one-third of the designs (options numbered 1-2, 7-8, 13-14, 19-20, 25-26, 31-32) employ "race" and "origin" terms to guide respondents to answer the question, "*What is Person 1's race or origin?*" These options are outlined in red. Another third of the designs (options numbered 3-4, 9-10, 15-16, 21-22, 27-28, 33-34) employ "race" and "ethnicity" terms to guide respondents to answer the question, "*What is Person 1's race or ethnicity?*" These options are outlined in yellow. The last third of the designs (options numbered 5-6, 11-12, 17-18, 23-24, 29-30, 35-36) employ no terms at all and use the general approach with the word "categories" to guide respondents to answer the question, "*Which categories describe Person 1?*" These options are outlined in green.



These various options are being tested to determine whether we can improve the understanding of the question concept and reduce confusion among respondents by using different terms (or no terms at all) for the race and ethnicity questions, such that we find more accurate and more reliable reporting between the survey and the reinterview Based on the research findings, a recommendation will be made for employing one of the options for terminology (race/origin vs. race/ethnicity vs. no terms at all – "categories").

3.7.6 | Developing a Recommendation for Best Question Design

Through all of these analyses, a recommendation on the best design for collecting and producing data on race/ethnicity for the 2020 Census will be made based on results from the 2015 NCT. The principal factor for making these decisions will be the web-based designs, as they are primary mechanism for a re-engineered 2020 Census. At the same time, we will examine the related factors for making decisions for the paper design for 2020. We will not choose different options that directly conflict with the design recommendations. For example, if a combined question is recommended, we will use it for both the web-based collections and paper data collections. Similarly, we will either employ a MENA category across the board, or it will not be used. These major design decisions will not differ by mode. Following this premise, we will use the same instruction wording and terminology concepts across modes (race/origin, race/ethnicity, or no terms at all – "categories").

We have a fully factorial approach for testing the different dimensions, as described earlier in Figure 3. Pooling together the treatment dimensions of the research, we are able to gain the power to analyze each of the dimensions focusing on the objectives of the NCT research on race and ethnicity (separate vs. combined; MENA vs. No MENA; instructions; terminology). Each of the 36 options we just described is shown in the graphic on the left, below, with a centrally placed number (from 1-36). The graphic on the right shows the corresponding options 1-36, surrounded by a quadrant of four colored blocks. The color shown for each of the four quadrants (with either a red block, yellow block, or green block) depicts the particular dimension being tested in each option. The upper left quadrant represents Dimension 1 (separate vs. combined); the upper right quadrant represents Dimension 2 (No MENA vs. MENA); the lower left quadrant represents Dimension 3 (old instructions vs. new instructions); and the lower right quadrant represents Dimension 4 (race/origin terms vs. race/ethnicity terms vs. no terms at all - "categories"). While this may look complex, it is a logical sequence of the different dimensions, and it enables us to put together the research dimensions to gain power and address the research questions.

3.7.7 | Explanation of Option Quadrants

We have provided a few examples to illustrate how this operates conceptually, and to explain how to understand the different components of the quadrants surrounding each of the options. Again, the color shown for each of the four quadrants (with either a red block, yellow block, or green block) depicts the particular dimension being tested in each option.

For example, each of the quadrants for Option 1 is shaded as red (see below). This indicates that the dimensions being tested are the separate question (red), no MENA category (red), old instructions (red), and race/origin terms (red).

OPTION 1	1	separate question **(red)**	no MENA category **(red)**
		old instructions **(red)**	race/origin terms **(red)**

Another example (see below) shows the quadrants for Option 20 are shaded yellow, green, green, and red. This indicates that the dimensions being tested are the combined question with write-in areas (yellow), MENA Category (green), new instructions (green), and race/origin terms (red).

OPTION 20	20	combined with write-in areas **(yellow)**	MENA category **(green)**
		new instructions **(green)**	race/origin terms **(red)**

A third example (see below) shows that all of the quadrants for Option 36 are shaded green.
This indicates that the dimensions being tested are the combined question with detailed checkboxes (green), MENA category (green), new instructions (green), and no terms – "categories" (green).

OPTION 36	36	combined with detailed checkboxes **(green)**	MENA category **(green)**
		new instructions **(green)**	no terms – "categories" **(green)**

3.7.8 | Examples of Design Decision Paths

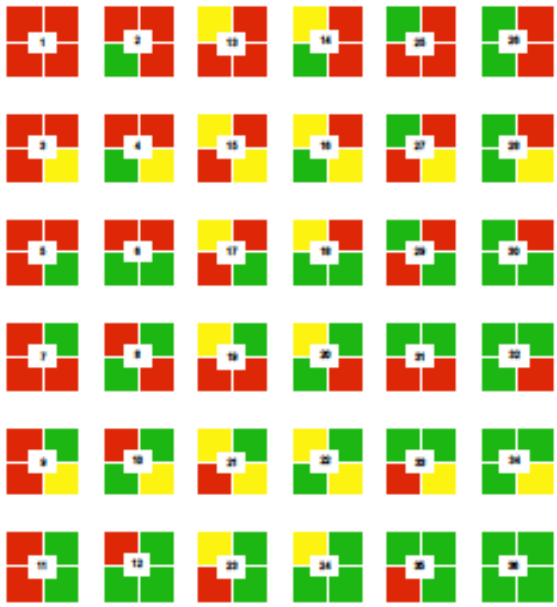
The following examples are provided to illustrate hypothetical scenarios for different design decision paths. Again, there are 36 different options that are being tested across the research dimensions. As each decision is made for the different research dimensions (e.g., separate vs. combined), the question designs which contain the element that is recommended will move forward in the decision process, and the question designs which contain the element that is not recommended will be dropped. To illustrate this step-by-step process, we present the following three hypothetical scenarios to show how different decisions would lead to one of the different options (1-36).

In a first hypothetical scenario, the results from the 2015 NCT research show that:

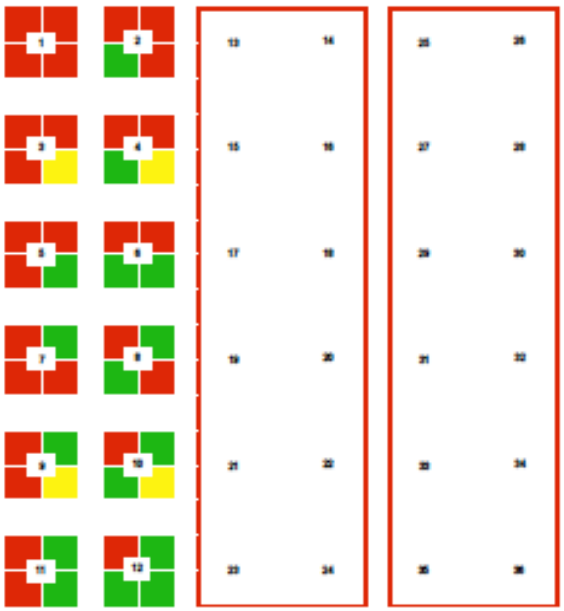
- the separate question approach performs better than either of the combined question approaches (with write-in areas; with detailed checkboxes)
- the question designs without a MENA category result in better data than the versions when a distinct MENA Category is included
- the old instruction to "Mark [***X***] one or more boxes" performs better than the new instructions
- the terms "Race" and "Origin" work better than other terminology

Following this scenario #1, we illustrate the step-by-step decisions through this process to reach the endpoint of a particular design option (number 1 through 36).

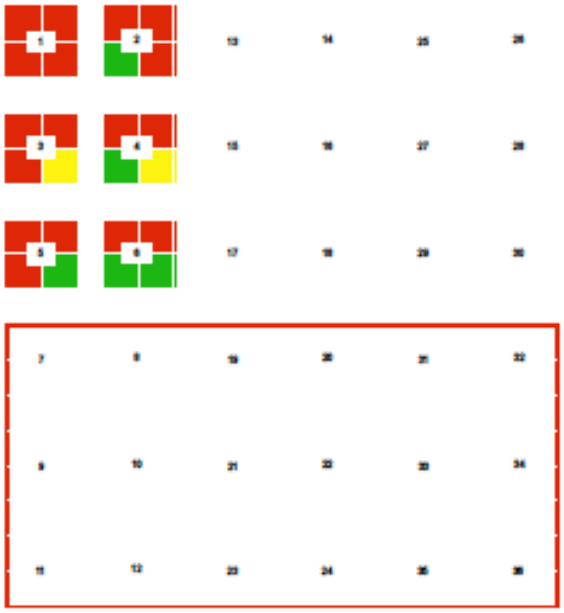
1. We begin with all 36 of the different options on the table:



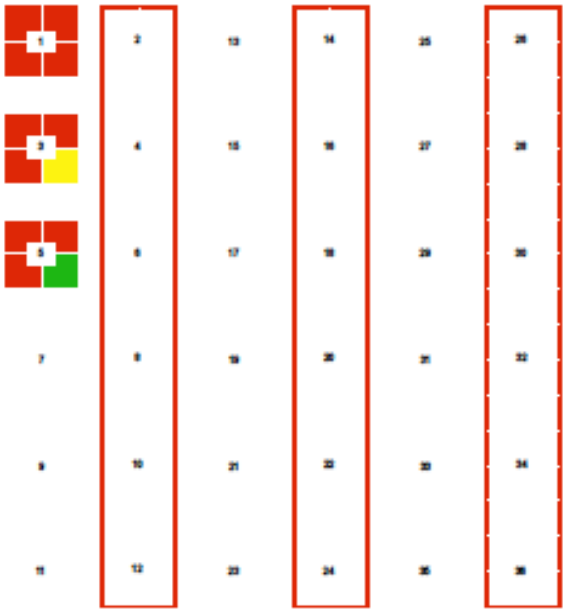
2. Next, the separate question recommendation moves forward, and both of the combined question approaches are dropped:



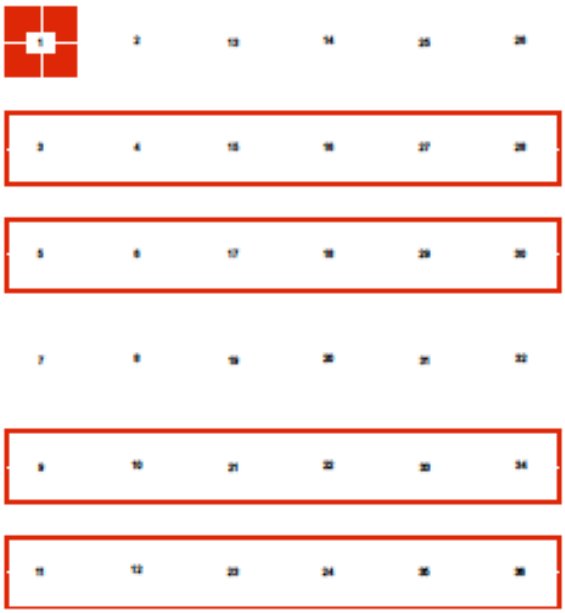
3. After that, we consider the MENA category. With the recommend to NOT use a distinct MENA category, we drop the options which contain a MENA category:



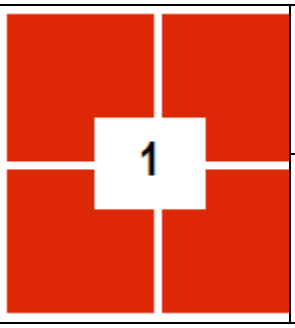
4. Next, different instructions are considered. The OLD instructions ("*Mark* [***X***] *one or more boxes*") are recommended to move forward, so we drop approaches with new instructions:



5. Finally, the "race" and "origin" terms are recommended to move forward, so the approaches with other terms are dropped:



6. Therefore, the design chosen for hypothetical scenario #1 is option 1 –
with the separate question (red),
no distinct MENA category (red),
old instructions (red),
and the use of race/origin terms (red).

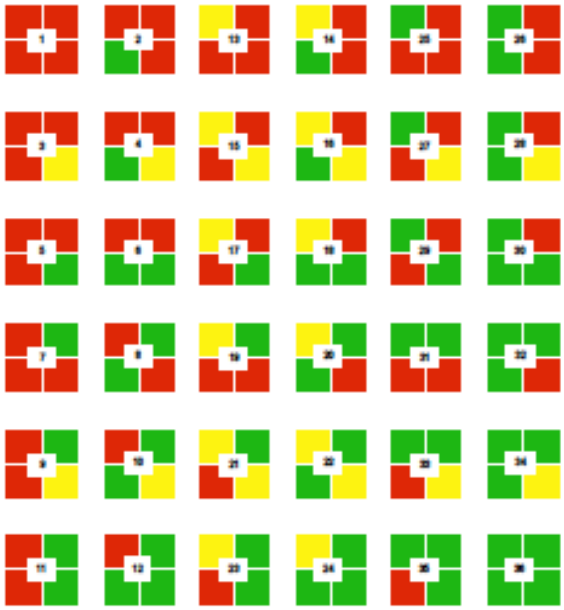
OPTION 1		separate question **(red)**	no MENA category **(red)**
		old instructions **(red)**	race/origin terms **(red)**

In a second hypothetical scenario, the results from the 2015 NCT research show that:

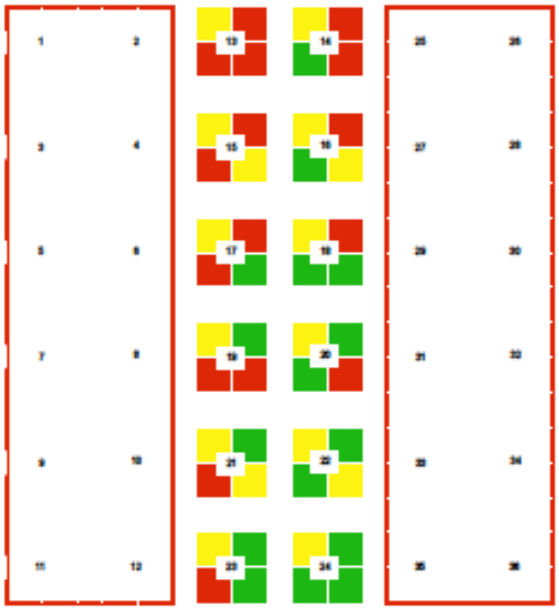
- the combined question approach with write-in areas performs better than the separate questions approach, and better than the combined question approach with detailed checkboxes
- having a distinct MENA category results in better data than versions without a MENA category
- new instructions to "*Mark all boxes that apply*" perform better than the old instructions
- the terms "race" and "origin" work better than other terminology

Following this scenario #2, we illustrate the step-by-step decisions through this process to reach the endpoint of a particular design option (number 1 through 36).

1. We begin with all 36 of the different options on the table:



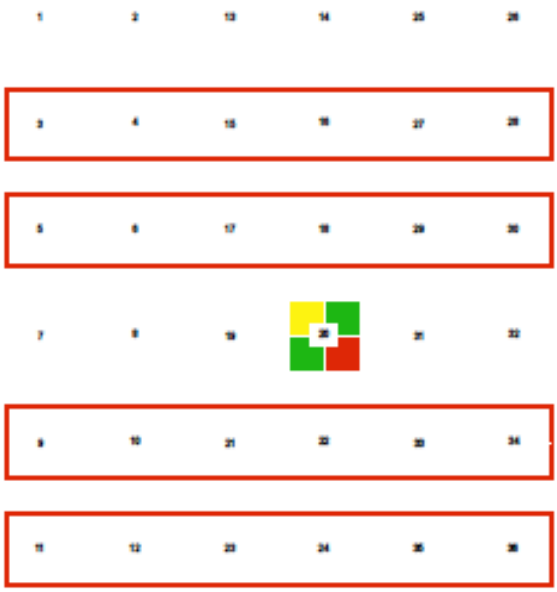
2. Next, the combined question with write-in areas is recommended, so we drop the separate questions approach and we drop combined question with detailed checkboxes approach:



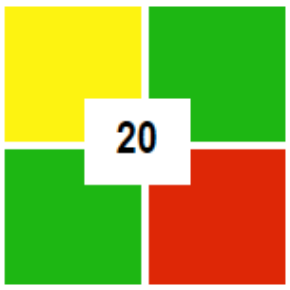
3. The designs with a distinct MENA category is recommended, rather than designs which do not include a MENA category:

4. New instructions ("*Mark all boxes that apply*") are recommended, so the approaches with old instructions are dropped:

5. The "race" and "origin" terms are recommended to move forward, so the approaches with other terms are dropped:



6. Therefore, the design chosen for hypothetical scenario #2 is option 20 – with the combined question with write-in areas (yellow), a distinct MENA Category (green), new instructions (green), and the race/origin terms (red).

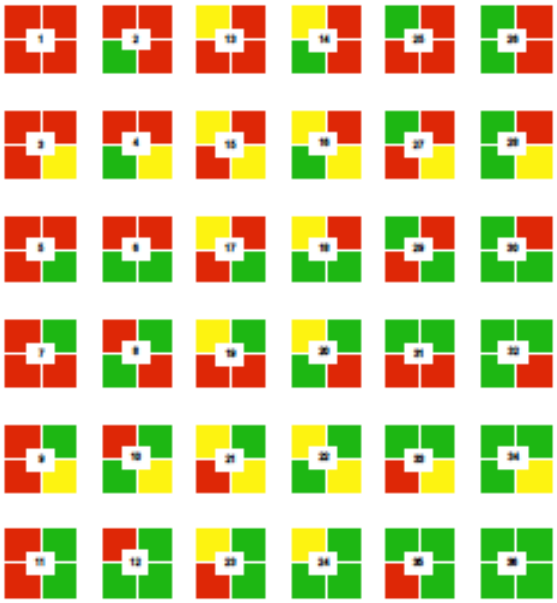
OPTION 20		Combined with write-in areas **(yellow)**	MENA category **(green)**
		New instructions **(green)**	race/origin terms **(red)**

In a third hypothetical scenario, the results from the 2015 NCT research show that:

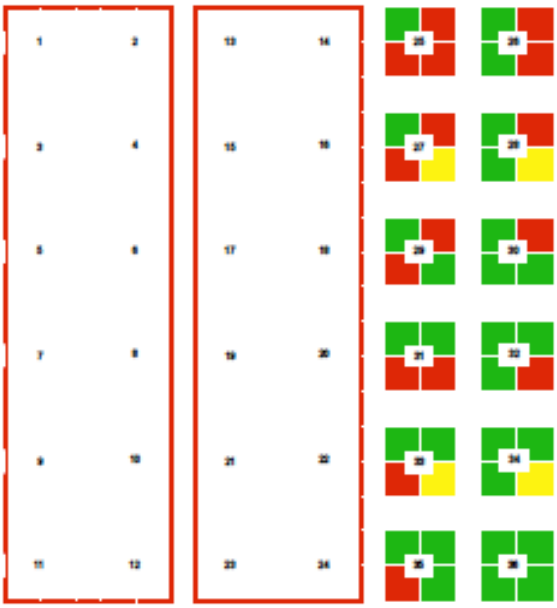
- the combined question approach with detailed checkboxes performs better than the separate questions approach, and better than the combined question approach with write-in areas
- having a distinct MENA category results in better data than versions without a MENA category
- new instructions to "*Mark all boxes that apply*" perform better than the old instructions
- the no terms approach (with "categories") works better than other terminology

Following this scenario #3, we illustrate the step-by-step decisions through this process to reach the endpoint of a particular design option (number 1 through 36).

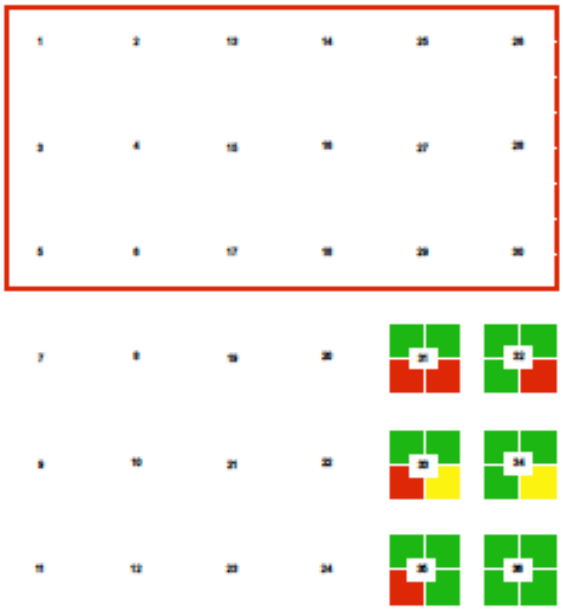
1. We begin with all 36 of the different options on the table:



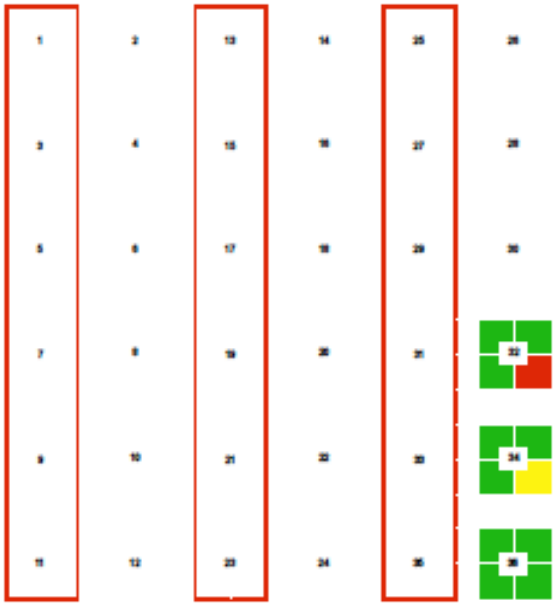
2. Next, the combined question approach with detailed checkboxes is recommended, so the separate question approach and the combined question with write-in areas approach are both dropped:



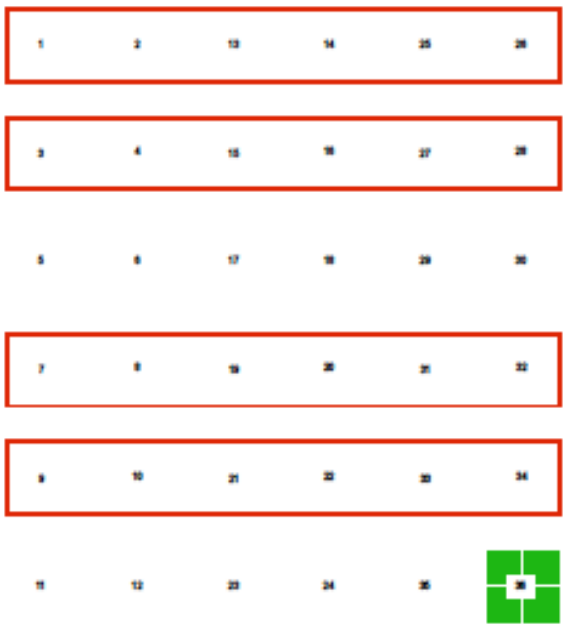
3. The designs with a distinct MENA category is recommended, rather than designs which do not include a MENA category:



4. The new instructions (“*Mark all boxes that apply*”) are recommended, so the approaches with old instructions are dropped:



5. The “race” and “ethnicity” terms are recommended to move forward, so the approaches with other terms are dropped:



6. Therefore, the design chosen for hypothetical scenario #3 is option 36 – with the combined question with detailed checkboxes (green), a distinct MENA Category (green), new instructions (green), and the no terms approach – with “categories” (green).

OPTION 36	36	combined with detailed checkboxes **(green)**	MENA category **(green)**
		new instructions **(green)**	no terms – “categories” **(green)**

Again, these three scenarios are hypothetical, and they are included to help to illustrate for readers the potential ways in which the different design decision paths may unfold. As part of our analytical research, we will examine all of the dimensions across the 36 different options that are being tested to determine which approaches are successful. As each decision is made for the different research dimensions (e.g., separate vs. combined), the question designs which contain the element that is recommended will move forward in the decision process, and the question designs which contain the element that is not recommended will be dropped. Ultimately, this will illustrate how the different decisions lead to one option as a final outcome and recommendation.

4. LIMITATIONS

Limitations to the 2015 NCT research include the following:

- Not being conducted within a decennial census environment
- Nonresponse bias
- Reinterview "truth"
- Statistical power for analyzing very small race and ethnic groups
- Possible reinterview conditioning effects
- Non-ID Internet respondents all get the same content

5. MILESTONE SCHEDULE

2015 NCT ACTIVITY	DATES
Data collection period	Monday, August 2014, 2015 - Saturday, October 30, 2015
Census Day	Tuesday, September 1, 2015
Reinterview period	Monday, September 21, 2015 - Monday, December 14, 2015
Draft study plan to critical reviewers	Wednesday, October 21, 2015
Residual coding of write-ins	Friday, December 4, 2015 - Wednesday, December 30, 2015
Study plan to Decennial Leadership Group	Friday, December 18, 2015
Study plan to Executive Steering Committee	Tuesday, January 12, 2016
Conduct data analysis	January - May, 2016 (planned)
Draft report to critical reviewers	Friday June 17, 2016 (planned)

6. DOCUMENT LOGS AND VERSION HISTORY

Verification of Document Content			
This document does not contain any: • Title 5, Title 13, Title 26, or Title 42 protected information; • Procurement information; • Budgetary information; and/or, • Personally identifiable information.			
Document Author/Team Lead:	*Nicholas A. Jones*	Date:	*2/10/2016*

The document version history recorded in this section provides the revision number, the version number, the date it was issued, and a brief description of the changes since the previous release. Baseline releases are also noted.

Version	Date	Description
1	10/21/2015	Initial draft of study plan submitted to Census Bureau critical reviewers.
2	12/18/2015	Updated draft incorporating revisions from Census Bureau critical reviewer comments. Study plan submitted to Decennial Leadership Group (DLG).
3	2/1/2016	Updated draft incorporating revisions from DLG comments and feedback from meeting with Census Bureau Executive Steering Committee (ESC). Prepared study plan for submission to key advisors (e.g., OMB Interagency Working Group on Race and Ethnic Research).
4	2/10/2016	Updated draft incorporating revisions from OMB comments on version 3. Submitted study plan to OMB Interagency Working Group on Race and Ethnic Research for review.

7. REFERENCES

Alberti, N. (2006). 2005 National Census Test: Analysis of the Race and Ethnicity Questions. DSSD 2005 Census Test Evaluations Memorandum Series #E-8, U.S. Census Bureau.

Childs, J., Terry, R., Jurgenson, N., Clifton, M., and G. Higbie. (2010). Iterative Cognitive Testing of the 2010 Race and Hispanic Origin Alternative Questionnaire Experiment (AQE) Reinterview. Statistical Research Division Research Report Series (Survey Methodology #2010-13), U.S. Census Bureau.

Compton, E., Bentley, M., Ennis, S., and Rastogi, S. (2012). 2010 Race and Hispanic Origin Alternative Questionnaire Experiment. Decennial Statistical Studies Division and Population Division, U.S. Census Bureau.

Dowling, J. (2014). *Mexican Americans and the Question of Race*. Austin, TX: University of Texas Press.

Dusch, G. (2011). 2010 Alternative Questionnaire Experiment Reinterview Instrument Specifications. Decennial Statistical Studies Division, 2010 Decennial Census Memorandum Series #B-13, 2010, U.S. Census Bureau.

Fernández, L., Gerber, E., Clifton, M., Higbie, G., and M. Meyers. (2009). Cognitive Pretesting of 2010 AQE Race and Hispanic Origin Treatment Panel. Statistical Research Division Research Report Series (Survey Methodology #2009-08), U.S. Census Bureau.

Hill, J. and M. Bentley (2014). 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment Supplemental Analysis: Race Distributions by Hispanic Origin. DSSD 2010 Decennial Census Memorandum Series #O-B-14-R3, U.S. Census Bureau.

Humes, K. and H. Hogan. (2009). "Measurement of Race and Ethnicity in a Changing, Multicultural America." *Journal of Race and Social Problems*. Volume 1, Issue 3.

Mathews, K. (2015). "Sample Design Specifications for the 2015 National Content Test." DSSD 2020 Decennial Census R&T Memorandum Series #R-11, U.S. Census Bureau.

Newby, C. and J. Dowling (2007). "Black and Hispanic: The Racial Identification of Afro-Cuban Immigrants in the Southwest." *Sociological Perspectives* 50.3: 343-66.

Rios, M., Romero, F, and Ramirez, R. (2014). "Race Reporting Among Hispanics: 2010." U.S. Census Bureau Working Paper Series Number 102.

Roth, W. (2012). *Race Migrations: Latinos and the Cultural Transformation of Race*. Standford, CA: Standford University Press.

Sheppard, D. W., Bentley, M., and Woltman, H. F. (2004). 2003 National Census Test: Analysis of Alternative Race and Ethnicity Questions Based on Panel Comparisons. Planning, Research, and Evaluation Division TXE 2010 Memorandum Series #DCO-NCT-F-05, U.S. Census Bureau.

U.S. Census Bureau (1997). Results of the 1996 Race and Ethnic Targeted Test, Population Division Working Paper No. 18.