

2005 National Census Test:
Analysis of the Race and Ethnicity Questions

Nicholas Alberti
Distributed for the November 30, 2006
Special Joint Meeting of the U.S. Census Bureau Advisory Committees

U S C E N S U S B U R E A U

**2005 NATIONAL CENSUS TEST:
ANALYSIS OF THE RACE AND
ETHNICITY QUESTIONS**

Nicholas Alberti
Decennial Statistical Studies Division

CONTENTS

EXECUTIVE SUMMARY	iii
1. INTRODUCTION	1
2. METHODOLOGY	3
2.1 Experimental Treatments	3
2.2 Sample Design	4
2.3 Estimation	5
2.4 Responses Based on Check Boxes	5
2.5 Hispanic Origin and Race Responses Were Not Edited	6
2.6 Self-Response Rates	6
3. LIMITATIONS	7
3.1 Results Reflect National Estimates for Only Self-response Mail Return Respondents	7
3.2 The 2005 NCT Cannot Make Inferences About Small Population Groups	7
3.3 Design Was Not Optimal for Testing Differences Between the Control Panel and Panel 6	7
3.4 Cases are Not Included for Areas Effected by Hurricane Katrina	8
4. RESULTS	8
4.1 Analysis of Three-Question Format	8
4.2 Analysis of Design Effects on Traditional Two-Question Format	17
4.3 Analysis of Question Design Effects Within the Three-Question Format	26
5. REFERENCES	29
Appendix A Examples of Hispanic Origin and Race Questions	31
Appendix B Summary of Test Panels and Their Treatments	37
Appendix C Table 15. Race Distributions for Hispanics (Control Panel & Panel 6) and Table 16. Race Distribution for Non-Hispanics (Control Panel & Panel 6)	38
Appendix D 2005 National Census Test Overview	40

LISTING OF TABLES

Table 1.	Self-Response Rates by Sampling Stratum	6
Table 2.	Self-Response Rates by Panel	7
Table 3.	Hispanic Origin Question Response Rates (Control Panel & Panels 2-5)	9
Table 4.	Race Question Response Rates (Control Panel & Panels 2-5)	10
Table 5.	Percent of Respondents That are Hispanic (Control Panel & Panels 2-5)	11
Table 6.	Percent of Hispanic Respondents Who Reported A Specific Hispanic Origin Group (Control Panel & Panels 2-5)	12
Table 7.	Race Distributions for Overall Population (Control Panel & Panel 6)	14
Table 8.	Percent of Respondents Within Major Race Groups Who Reported a Detailed Race Group (Control Panel & Panel 6)	15
Table 9.	Hispanic Origin Question Response Rates (Control Panel & Panel 6)	19
Table 10.	Race Question Response Rates (Control Panel & Panel 6)	19
Table 11.	Percent of Respondents That are Hispanic (Control Panel & Panel 6)	20
Table 12.	Percent of Hispanic Respondents Who Reported A Specific Hispanic Origin (Control Panel & Panel 6)	21
Table 13.	Race Distributions (Control Panel & Panel 6)	22
Table 14.	Percent of Respondents Within Major Race Groups Who Reported a Detailed Race Group (Control Panel & Panel 6)	24
Table 15.	Race Distributions for Hispanics (Control Panel & Panel 6)	38
Table 16.	Race Distribution for Non-Hispanics (Control Panel & Panel 6)	39

EXECUTIVE SUMMARY

As part of the Census Bureau's 2010 research and testing program on race and ethnicity, the U.S. Census Bureau conducted the 2005 National Census Test (NCT) during the summer of 2005. It is one in a series of tests that will allow the Census Bureau to finalize content, methodology, and operational procedures in time to conduct a 2008 Census Dress Rehearsal. The purpose of the 2005 NCT race and ethnicity component was to improve the completeness and accuracy of Hispanic origin and race data through improved wording, instructions, examples and form design; and to increase self reporting by respondents into the Hispanic origin and race categories identified by the Office of Management and Budget (OMB) and used by the federal statistical system. The 2005 NCT race and ethnicity sample was about 180,000 housing units selected from Census 2000 mailback areas. It consisted of five experimental panels and a control panel with 30,000 housing units in each.

The major focus of the 2005 NCT race and ethnicity component was to 1) evaluate an improved Census 2000 two-part Hispanic origin and race question format (Panel 6), and 2) evaluate a new three-question format that included simplified versions of the Hispanic origin and race questions and a separate ancestry question (Panels 2 through 5). The question on ancestry was intended to elicit detailed information on all race and ethnic groups. The Census Bureau's overall goal is to reduce the number of missing Hispanic origin and race responses.

The test included six short form panels to test changes to the Hispanic origin and race questions. A modified Census 2000 two-question format was used as the control against which the outcomes for the other five panels were compared. The test Panel 6 also used a two-question format to collect data on race and ethnicity. As in the Control Panel, the Hispanic origin and race questions included extensive lists of specific Hispanic origin check boxes and detailed race group check boxes. Some of the differences between Panel 6 and the Control Panel are: 1) both included similar but slightly different instructions to answer both the Hispanic origin and race questions; 2) the statement 'For this Census Hispanic origins are not races' was included in Panel 6 but not the Control Panel; 3) an instruction in the Control Panel to 'Mark "No" if not of Spanish, Hispanic, or Latino origin' was not included in Panel 6; 4) the Hispanic origin check box responses were double banked (two columns) in the Control Panel but single banked (one column) in Panel 6; and 5) the instruction to the race question in the Control Panel included the phrase 'to indicate what this person considers himself/herself to be' but this was not included in Panel 6.

Four experimental panels (Panels 2-5) used an alternative three-question format with simplified Hispanic origin and race questions and a question on ancestry. It was hypothesized that this format would improve self-response, and the ancestry question could collect detailed data for all groups.

The simplified Hispanic origin question in the three-question format of Panels 2-5 included two check box answers, Yes and No, which allowed respondents to report that they are or are not

Hispanic. The race question in Panels 2-5 included check boxes for five federally recognized race categories -- American Indian or Alaska Native (AIAN), Asian, Black or African American, Native Hawaiian or Other Pacific Islander (NHPI), and White -- plus a sixth category for Some Other Race (SOR). Two versions of the ancestry question were included in Panels 2-5. Both required a write-in response and included examples of specific Hispanic origin groups and detailed race groups. One provided more explanation of what ancestry means and the other basically instructed the respondent to report his or her ancestry. There were minor wording differences in questions and instructions across Panels 2-5. These differences were included to test for the best options in the context of the three-question format. In addition, Panel 5 included a separate question for AIAN respondents to report their tribe.

Results of Testing the Three-Question Format

The three-question format (Panels 2-5) resulted in some loss of completeness of data about Hispanic origin.

The information on Hispanic origin is much less complete in the three-question format (Panels 2-5). In the Control Panel, 93.52 percent of Hispanic respondents reported a specific Hispanic origin group. But in Panels 2-5, the percentages of Hispanics who reported a specific Hispanic origin group are at least thirty percentage points lower than in the Control Panel. Examples of specific Hispanic origin groups are Mexican, Cuban, and Puerto Rican.

The percent of Hispanics is slightly higher in the three-question format: 9.02 percent in the Control Panel and between 10.00 percent and 10.35 percent in Panels 2-5. But as mentioned above, Hispanics provided less detailed data about Hispanic origin in Panels 2-5 than in the Control Panel.

The results on the level of response to the Hispanic origin question are mixed. While response to the Hispanic origin question is high on all panels, ranging from 96.39 percent to 97.41 percent, the Hispanic origin response rates for Panels 3 and 4 are higher than in the Control Panel by an estimated 0.76 and 0.64 percentage points, respectively. The response rates of Panels 2 and 5 are not statistically different from the response rates of the Control Panel.

The three-question format (Panels 2-5) produced a higher response to the race question for some respondents but the race distributions differ from the Control Panel.

For the overall population, the response rate to the race question is high in all panels and no differences in the response rate to the race question were found between Panels 2-5 and the Control Panel. The response rates to the race question for all respondents ranged from 95.92 percent to 96.50 percent across the five panels (the Control Panel and Panels 2-5). For Hispanics, the response to the race question differed between the Control Panel and three of the alternative panels (Panels 3, 4 and 5). The race question response rates of Hispanics in these three panels were less than five percentage points higher than the Control Panel. The race

question response rate of Hispanics was 69.48 percent in the Control Panel and ranged from 71.92 percent to 74.39 percent across Panels 2-5.

The interpretation of these results is complicated by the fact there were different race distributions in Panels 2-5 compared with those from the Control Panel. There appears to be a shift from reporting race as 'White' to reporting race as 'Some Other Race' in Panels 2-5. Fewer persons reported their race as 'White' than do those from the Control Panel, while about twice as many respondents reported their race as 'Some Other Race' in Panels 2-5. The sum of the percentages of respondents who reported their race as 'White' and 'Some Other Race' in Panels 2-5 are not very dissimilar from the corresponding sum of percentages in the Control Panel. The sum of the percentages in Panels 2-5 ranged from 80.90 percent to 81.24 percent compared with 81.05 percent for the Control Panel.

Improving data for use in the Census Bureau's Intercensal Population Estimates Program was an objective of introducing the three-question format which has a separate question on ancestry. The Intercensal Population Estimates Program and other federal statistical programs that rely on that data do not include an SOR race group category. Census Bureau must impute one of the five OMB recognized races to SOR respondents in order to meet the requirements of those programs. The race data collected on the three-question format cannot be used as effectively by the Intercensal Population Estimates Program compared with the race data from the traditional two-question format. This is evident from the percentages of SOR respondents on the Control Panel (1.96 percent) compared with Panels 2-5 (3.71 percent to 4.74 percent). A higher percent of respondents reported race as SOR in each of Panels 2-5 compared with the Control Panel. The percentages of Hispanic SOR respondents are particularly high on Panels 2-5 (26.73 percent to 36.51 percent).

The three-question format (Panels 2-5) resulted in a loss in the completeness of data for Asians and NHPI.

Compared with the Control Panel, the percentages of Asians and NHPI who reported a detailed race group are lower in Panels 2-5. In the Control Panel, 92.79 percent of Asians reported a detailed race group, while the percentages for Panels 2-5 ranged from 82.64 percent to 84.08 percent. The differences between the Control Panel and each of Panels 2-5 are particularly large for NHPI respondents. In the Control Panel, 96.28 percent of NHPI respondents reported a detailed race group (e.g., Guamanian, Samoan). In Panels 2-5, the percent of NHPI who reported a detailed race group ranged from 30.01 percent to 58.19 percent.

Results of Testing the Two-Question Format

Data on Hispanic origin are more complete in Panel 6.

The response to the Hispanic origin question was high on both the Control Panel (96.65) and Panel 6 (95.00). A higher percent of respondents reported they are Hispanic in Panel 6 (9.72 percent) than in the Control Panel (9.02 percent). Because the percent of Hispanics was higher in Panel 6 but the response to the Hispanic question was lower in Panel 6, we know that non-Hispanics skipped the Hispanic origin question in Panel 6 relatively more often than Hispanics did. This shows that non-Hispanics contributed more to the lower response in Panel 6 than did Hispanics.

More complete Hispanic origin data were collected in Panel 6 compared with the Control Panel. The percent of Hispanics who reported a specific Hispanic origin was 96.07 percent in Panel 6 and 93.52 percent in the Control Panel.

It is difficult to know what features of the two versions of the Hispanic origin question influenced these results, since there were multiple differences between the two questionnaires with respect to the wording and the format. The phrase “Mark ‘No’ if not of Spanish, Hispanic or Latino origin” was included in the Control Panel but not in Panel 6, and the order in which the Hispanic identifiers (Spanish, Hispanic, and Latino) were listed differs. Also, the Hispanic origin question check box response categories are double banked in the Control Panel but single banked in Panel 6. It is not possible to determine how much each of these differences contributed to the results independently or in combination. However, the results clearly point to more complete Hispanic origin data for respondents in Panel 6.

Both the Control Panel and Panel 6 perform equally well with respect to race data, both for Hispanics and non-Hispanics.

No statistical differences were found in the response rates to the race question between the Control Panel and Panel 6. Also, there was little difference in the completeness of the race data. With the exception of the NHPI race group, there were no statistical differences between the panels in the percent who reported a detailed AIAN, Asian, or SOR race group. NHPI respondents reported a detailed race group 96.28 percent of the time in the Control Panel and 73.00 percent of the time in Panel 6.

1. INTRODUCTION

As part of the Census Bureau's 2010 research and testing program on race and ethnicity, the U.S. Census Bureau conducted the 2005 National Census Test (NCT) during the summer of 2005. It is one in a series of tests that will allow the Census Bureau to finalize content, methodology, and operational procedures in time to conduct a 2008 Census Dress Rehearsal. The purpose of the 2005 NCT race and ethnicity component was to improve the completeness and accuracy of Hispanic origin and race data through improved wording, instructions, examples and form design; and to increase self reporting by respondents into the Hispanic origin and race categories identified by the Office of Management and Budget (OMB) and used by the federal statistical system. The 2005 NCT race and ethnicity sample was about 180,000 housing units selected from Census 2000 mailback areas. It consisted of five experimental panels and a control panel with 30,000 housing units in each.

The 1997 OMB standards for collecting, tabulating, and presenting data on race and ethnicity are the basis for the Census Bureau's procedures. These standards identify five federally recognized race categories: American Indian or Alaska Native (AIAN), Asian, Black or African American, Native Hawaiian or Other Pacific Islander (NHPI), and White. With the OMB's approval, the Census Bureau has included a sixth racial category, Some Other Race (SOR). The procedures allow respondents to report one or more races. The standards also identify two ethnicity categories: Hispanic or Latino, and Not Hispanic or Latino. The Census Bureau places a question on Hispanic origin before a separate question on race, which is in line with the OMB's recommendations on the number and placement of these questions when self-identification is used to report race and ethnicity.

The 2005 NCT research builds on the research of the 2003 National Census Test (NCT). Alternative wording and response categories for the Hispanic origin and race questions were tested in the 2003 NCT. The 2003 NCT used the traditional two-question format to collect race and ethnicity data. That format uses one question for Hispanic origin and one question for race. Each of these questions included extensive check box responses. The wording alternatives tested in 2003 included the use of examples, the inclusion of the term "origin" in the Hispanic origin question and modification to the respondent instructions. The 2003 NCT results indicate that the changes reduced the reporting of generic response such (as "Hispanic," "Latino," or "Spanish") (Martin, et al., 2004). The Hispanic origin questions tested in the 2005 NCT incorporated those wording changes.

The effect of adding examples to questions was also tested in the 2003 NCT by including examples in the "Other Spanish, Hispanic, Latino", "Other Asian", and "Other Pacific Islander" response categories, in the respective questions. The results were mixed but showed that the inclusion of examples tended to reduce the reporting of generic Hispanic origin and race responses (Martin et al., 2004). The 2005 NCT questionnaires that used the traditional two-question format to collect data on race and ethnicity included examples in a fashion similar to what was tested in the 2003 NCT. The inclusion of examples was further tested in the 2005 NCT

with respect to an experimental three-question format which is a departure from the two-question format for collecting data on race and ethnicity.

The 2005 NCT also continued testing a modified respondent instruction that encourages respondents to answer both questions. The test of the modified instruction in the 2003 NCT showed mixed results. The instruction resulted in fewer Hispanics reporting as ‘Some Other Race’ and increased their response to race. However, it increased item nonresponse to the Hispanic origin question, and for non-Hispanics, it increased the item nonresponse to race.

The 2005 NCT tested an alternative three-question format for collecting data on race and ethnicity. This format used a simplified Hispanic origin and race questions that were shorter than those used in the traditional two-question format. The three-question format was suggested as a method to increase self-identification without losing detailed information about Hispanic origin and race. It was hypothesized that the simplified Hispanic origin and race questions would improve self-response, and that the ancestry question could collect detailed race and ethnicity data for all groups.

The simplified Hispanic origin question included only “Yes” and “No” response choices. The simplified race question included six race response boxes for the five race groups -- American Indian or Alaska Native (AIAN), Asian, Black or African American, Native Hawaiian or Other Pacific Islander (NHPI), and White -- plus Some Other Race (SOR). The experimental three-question format also included a question on ancestry. This question requires a write-in response and is used to capture information on specific Hispanic origin groups and detailed race groups.

2005 NCT Race and Ethnicity Test Objectives

Five broad objectives were set for testing of Hispanic origin and race questions in the 2005 NCT. These are:

1. Maximize self-identification by reducing nonresponse.
2. Shorten Hispanic origin and race questions and get detailed information for all groups.
3. Obtain information on tribal affiliation and enrollment.
4. Assess the use of ancestry data for editing and imputing missing data for Hispanic origin and race.
5. Improve the utility of the Hispanic origin and race data for the Intercensal Population Estimates Program.

2. METHODOLOGY

2.1 Experimental Treatments

The 2005 NCT included six sample design panels to test the alternative treatments for collecting race and ethnicity data. The questionnaires for two of these panels (the Control Panel and Panel 6) used a traditional two-question format for collecting these data. The questionnaires for the four other panels (Panels 2-5) used an experimental three-question format which had simplified questions on Hispanic origin and race, and a separate question on ancestry. Illustrations of the race and ethnicity questions for these panels are included in Appendix A. A summary of the experimental treatments by panel is shown in Appendix B.

Control Panel and Panel 6

The Control Panel and Panel 6 both used the traditional two-question format with one question on Hispanic origin followed by one question on race. These two questions were preceded by a simple instruction to answer both questions. The Hispanic origin question included an extensive list of specific Hispanic origin group check boxes. The race question included an extensive list of major race group and detailed race group check boxes.

The differences between these two questionnaire panels are as follows:

- The wording of the initial respondent instruction for responding to the Hispanic origin and races questions differ between the two panels. On the Control Panel it is “Please answer BOTH Questions 8 and 9”. On Panel 6 the instruction is “Please answer BOTH Question 8 about Hispanic origin and Question 9 about race”.
- A note to the respondent that says “Hispanic origins are not races” follows the initial instruction to answer both questions on Panel 6, but the note is not included in the Control Panel.
- The Hispanic origin identifiers were listed in different orders in the Hispanic origin questions. The Control Panel lists them as Spanish, Hispanic or Latino origin, while Panel 6 lists them as Hispanic, Latino, or Spanish origin.
- The Hispanic origin question of Control Panel includes the instruction: ‘Mark “No” if not of Spanish, Hispanic or Latino origin’. This instruction is not included in the Panel 6.
- The race questions of these two panels included slightly different instructions. Panel 6 uses the simple instruction ‘Mark one or more boxes’. The instruction in the Control Panel is ‘Mark one or more races to indicate what this person considers himself/herself to be’.

- The check box responses for the Hispanic origin question were double banked (two columns) in the Control Panel but single banked (one column) in Panel 6.

Panels 2-5

Each of the Panels 2-5 used an experimental three-question format to capture race and ethnicity data. The first two questions asked about Hispanic origin and race. The third question asked about ancestry.

The Hispanic origin question included two check box answers, “Yes” and “No”, which allowed the respondent to report that they are or are not Hispanic. The race question included six check boxes; five for the federally recognized race categories -- American Indian or Alaska Native (AIAN), Asian, Black or African American, Native Hawaiian or Other Pacific Islander (NHPI), and White-- plus a sixth check box for Some Other Race (SOR).

The question on ancestry is included to capture detailed information on all race and ethnicity groups not captured in the simplified Hispanic origin and race questions. Examples are included in the ancestry question in all of these four panels.

Panel 3 is the simplest of the four panels that used the experimental three-question format. It uses the briefest questions and instructions. In relationship to Panel 3, the other three questionnaires differ in the following ways:

- Panel 2- Examples are included in the Hispanic origin and race question of this questionnaire. It is the same as Panel 3 in all other respects.
- Panel 4 - The race question included the additional instruction “ For this census, Hispanic origins are not races”. The ancestry question includes the added explanation that “People in the United States are from many countries, tribes, and cultural groups”.
- Panel 5 - In place of a single ancestry question, this panel also included a separate question for AIAN respondents to indicate their tribe(s) as well as whether or not they are enrolled in a tribe. The ancestry question was modified to reflect the inclusion of the separate question on AIAN tribe. The ancestry question includes the explanation that “People in the United States are from many countries and cultural groups”. Additionally the ancestry question included the check box response “Don’t Know” and “Only United States” for respondents who were unable to identify a specific ancestry.

2.2 Sample Design

The 2005 NCT was a multifaceted test that studied questionnaire content and design as well as methodology. The sample design included 24 panels, six of which were used to test

questionnaire content and design to improve the completeness and accuracy of race and ethnicity data.

The sample for the 2005 NCT was selected from census addresses in mailout areas. They are areas with city-style addresses and they included about 80 percent of all housing units in Census 2000. The sampling universe included all 50 states and the District of Columbia. It did not include the U.S. territories.

The sampling universe was stratified into two strata: 1) High Non-White or Hispanic concentration stratum, and 2) Low Non-White or Hispanic concentration stratum. The High Non-White or Hispanic concentration stratum was comprised of census blocks that had high concentration of Non-White or Hispanic populations. About 35 percent of the sample universe fell into the High Non-White or Hispanic concentration stratum.

A sample of 30,000 housing units was selected for each of the six panels used to test improvements to Hispanic origin and race data. Equal sample sizes of 15,000 housing units from each stratum were selected which resulted in the over sampling of the population in the High Non-White or Hispanic concentration stratum.

2.3 Estimation

All analysis metrics are based on sample estimates. Estimated standard errors for all estimates were computed using a stratified jackknife replication procedure. This computation method accounts for the stratification in the sample, which we expect to result in estimated standard errors that are lower than those for a simple random sample of equal size. Housing units were sorted into the same order in which they were selected and clusters of housing units (or housing units selected at each hit) were assigned sequentially to replicates. The use of this assignment approach also accounts for the clustering of people within a household in computing errors for person level estimates, since people within households are clustered together in the same replicate.

2.4 Responses Based on Check Boxes

Based on the initial data captured, the responses to the Hispanic origin and the race questions were based solely on the check boxes marked and response to each question was treated as missing if no check box is filled. For Panels 1 and 6, this meant that write-in responses to Hispanic origin and race questions were not treated as responses unless the associated check box was filled. In Panels 1 and 6, 0.17 percent and 0.26 percent of respondents, respectively, failed to check a response box but provided write-in response that was a specific Hispanic origin group. About 1.30 percent and 1.21 percent in Panels 1 and 6, respectively, failed to check a response but provided a detailed race group write-in response.

2.5 Hispanic Origin and Race Responses Were Not Edited

Write-in responses to Hispanic origin, race, and ancestry responses were not used to edit the check box responses. No Hispanic origin or race identified from the check box responses were altered based on additional information provided by a write-in response.

2.6 Self-Response Rates

The self-response rates are shown in Tables 1 and 2. The self-response rates include both mail returns and internet responses. Table 1 shows the mail response rate (percent response), the percent internet response, and the percent of questionnaires that were undeliverable as addressed (UAA). These values are shown for each sampling stratum across all 24 panels in the 2005 NCT. Table 2 shows the self-response rate for each of the six panels used to evaluate the race and ethnicity questions. The self-response rates were calculated as:

$$\text{Self-response rate} = \frac{\# \text{ of households with a nonblank return}}{\text{sample size} - \text{UAA addresses}} \times 100$$

Table 1
SELF-RESPONSE RATES
BY SAMPLING STRATUM
Percent

	Total	Mail Response	Internet Response	UAA
Full Sample	60.7	53.3	7.3	7.8
High Non-White and Hispanic Stratum	44.4	39.6	4.8	8.3
Low Non-White and Hispanic Stratum	68.1	59.6	8.5	7.5

Table 2
SELF-RESPONSE RATES
BY PANEL

Panel	Percent
Control	61.2
2	60.4
3	61.1
4	60.8
5	60.7
6	60.6

3. LIMITATIONS

3.1 Results Reflect National Estimates for Only Self-response Mail Return Respondents

Since the sample included only housing units in Census 2000 mailback areas and there was no follow-up to housing units that did not respond via mail results of the 2005 NCT cannot be generalized to the entire population. Past censuses have shown that mail response is lower for Hispanics and racial groups other than Whites. For example, in the Census 2000, the mail response was 69.5 percent for Hispanic households and 79.3 percent for non-Hispanic households. Additionally, the data were not adjusted for nonresponse. Thus, the estimates of people by Hispanic origin and race cannot be generalized to the national distributions.

3.2 The 2005 NCT Cannot Make Inferences About Small Population Groups

The 2005 NCT was not designed to allow the Census Bureau to make inferences about small population groups. This was a national mail test and did not include a content reinterview. Accordingly, while this evaluation reports certain population distributions, one should not use those distributions to make inferences about the number or percent of any population group other than those discussed within the report.

3.3 Multiple Differences Between the Control Panel and Panel 6 Confound the Analysis of Effects

The 2005 NCT was designed to test the overall effects of the Control Panel and Panel 6. It is not an optimal experimental design for testing differences between these panels in the Hispanic

origin questions or race questions. Inferences can be made only for the combined effect of the paired Hispanic origin and race questions.

Also, the 2005 NCT was not designed to test the effects of individual features within these questions. There are multiple differences between these panels in the wording and format of the Hispanic origin and race questions, and in related instructions. For example, differences between the Hispanic origin questions on these two panels were: 1) the Control Panel included the instruction to “Mark ‘No’ if not of Spanish, Hispanic, or Latino origin”, but this instruction was not included in Panel 6; 2) the check box responses were double banked in the Control Panel but single banked in Panel 6; and 3) the order of the Hispanic identifiers (i.e., Spanish, Hispanic, Latino) different in the two panels. The potential confounding of these features prevents drawing conclusions about the individual features.

3.4 Cases are Not Included for Areas Affected by Hurricane Katrina

Hurricane Katrina occurred on the Monday following the Friday evening post office delivery of the initial mailing of questionnaire packages. The postal service reported that mail delivery was suspended to areas that encompassed 3,200 cases. No replacement mailings were sent to addresses in these areas. Cases in these areas were treated as UAAs.

4. RESULTS

The analysis is divided into three broad topics. Section 4.1 present the results of the analysis of the effects of the experimental three-question format (Panel 2-5). Section 4.2 addresses the design effects of alternative question wording in the traditional two-question format (Control Panel and Panel 6). Section 4.3 shows the results of the analysis of the effects of question wording alternatives on three-question format.

4.1 Analysis of Three-Question Format

This section analyzes the effects of the three-question format that includes shorter questions on Hispanic origin and race and a separate question on ancestry. Four questionnaires with the three-question format were tested. These are represented by design Panels 2-5. These questionnaires differ from each other with respect to question wording, inclusion or exclusion of examples, and the format for collecting ancestry data.

The data on Hispanic origin and race from each of these panels are compared with the results from the Control Panel. The goal of this analysis is to identify which, if any, of the questionnaires in Panels 2-5 resulted in a higher response to Hispanic origin and race questions and more complete data on Hispanic origin and race compared to the Control Panel.

4.1.1 What is the effect on item response rates?

Response to the Hispanic origin question

Table 3 shows the response rates for the Hispanic origin question. Response to Hispanic origin was slightly higher on two of the three-question format panels, Panels 3 and 4. The questionnaires for these two panels are the simplest of the four panels in that neither included examples in the Hispanic origin and race question, and they do not include a separate question on AIAN tribe.

Table 3
HISPANIC ORIGIN QUESTION
RESPONSE RATES
(Control Panel & Panels 2-5)
Percent of All Respondents

Panel	
1 - Control Panel	96.65 (0.14)
2 - Short Format w/Examples	96.83 (0.12)
3 - Short Format	97.41* (0.11)
4 - Short Format w/Added Race Instruction	97.29* (0.11)
5 - Short Format w/Separate Tribe Response	96.39 (0.12)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.*

Response to the race question

Table 4 shows the response rates to the race question for all respondents, for Hispanics only, and for non-Hispanics only. In comparison with the Control Panel, Hispanics had a higher response rate for three of the three-question format panels. These panels were Panels 3, 4, and 5. The only three-question format panel that does not have a higher response the for Hispanics than the Control Panel is Panel 2. It is the only three-question format panel that includes examples in the race question.

Table 4
RACE QUESTION RESPONSE RATES
(Control Panel & Panels 2-5)
Percent of All Respondents by Hispanic Origin

Panel	All Respondents	Hispanics	Non-Hispanics
1 - Control	96.28 (0.17)	69.48 (1.36)	99.40 (0.05)
2 - Short Format w/Examples	95.92 (0.16)	71.92 (1.15)	99.28 (0.06)
3 - Short Format	96.50 (0.16)	73.62* (1.19)	99.47 (0.05)
4 - Short Format w/Added Race Instruction	96.37 (0.17)	73.35* (1.24)	99.44 (0.06)
5 - Short Format w/Separate Tribe Response	96.45 (0.15)	74.39* (1.12)	99.25* (0.06)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.*

4.1.2 What is the effect on the distribution of Hispanics and on the reporting of specific Hispanic origin groups?

Table 5 shows the percent of the entire respondent population that reported that they were Hispanic and the percent that provided a specific Hispanic origin (e.g., Mexican, Puerto Rican) in response to the Hispanic origin question. Table 6 shows the percent of Hispanics who provided a specific Hispanic origin.

Respondents could provide a specific Hispanic origin in two ways in the Control Panel, either by marking any of the three specific Hispanic origin check box responses (Mexican/ Mexican American/Chicano, Puerto Rican, and Cuban) or by marking the check box for “Yes, another Hispanic, Latino, or Spanish origin” and writing a specific Hispanic origin in the space provided. By contrast, in the Panels 2-5, the respondent needed to answer a separate ancestry question to report a specific Hispanic origin group. That ancestry question required a write-in response.

More respondents reported that they were Hispanic in Panels 2-5 than in the Control Panel (see Table 5). However, the results in Tables 5 and 6 show that data on specific Hispanic origin were less complete in Panels 2-5. As Table 5 shows, fewer respondents in the total population provided a specific Hispanic origin. Also, fewer Hispanics in Panels 2-5 provided a specific Hispanic origin compared with Hispanics in the Control Panel (Table 6).

Table 5
PERCENT OF RESPONDENTS THAT ARE HISPANIC¹
(Control Panel & Panels 2-5)
 Percent Among All Respondents by Panel

	Control	2	3	4	5
Hispanics	9.02 (0.26)	10.24* (0.27)	10.35* (0.27)	10.08* (0.27)	10.00* (0.26)
Hispanics Who Reported a Specific Hispanic Origin	8.44 (0.25)	6.12* (0.21)	6.44* (0.22)	6.35* (0.22)	6.35* (0.21)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.*

¹ Note that the percent of Hispanics in the population as shown in Table 5 is an estimate of the distribution of Hispanics in mail response households. Based on Census 2000 data, the expected percent for the entire U.S. population will be larger.

Table 6
PERCENT OF HISPANIC RESPONDENTS WHO REPORTED
A SPECIFIC HISPANIC ORIGIN GROUP
(Control Panel & Panels 2-5)

Percent of Hispanics by Panel

	Control	2	Panel 3	4	5
Respondents Who Did Not Report a Specific Origin	6.48 (0.65)	40.19 (1.30)	37.75 (1.31)	37.01 (1.28)	36.55 (1.21)
Respondents Who Reported a Specific Origin	93.52 (0.65)	59.81* (1.30)	62.25* (1.31)	62.99* (1.28)	63.45* (1.21)
<i>Specific Origin Groups</i> ²					
Mexican	59.25 (1.15)	27.36 (1.21)	28.97 (1.26)	32.48 (1.30)	29.38 (1.24)
Puerto Rican	11.70 (0.89)	7.16 (0.67)	6.90 (0.69)	8.10 (0.76)	7.32 (0.65)
Cuban	4.35 (0.49)	3.88 (0.52)	2.86 (0.42)	3.73 (0.47)	4.21 (0.53)
Dominican	1.34 (0.31)	1.89 (0.33)	2.71 (0.45)	2.14 (0.36)	2.47 (0.42)
Central American	5.54 (0.61)	5.83 (0.60)	5.94 (0.60)	4.71 (0.55)	5.50 (0.58)
South American	7.07 (0.78)	4.92 (0.61)	4.20 (0.52)	5.23 (0.60)	6.86 (0.69)
Latin American	0.02 (0.02)	0.25 (0.10)	0.65 (0.20)	0.23 (0.10)	0.46 (0.17)
Other Hispanic	4.26 (0.50)	8.52 (0.72)	10.01 (0.82)	6.37 (0.61)	7.25 (0.62)

* The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.

² The percentages in this section of the table sum to the total percentages for “People Who Reported a Specific Origin”. The percent for each specific Hispanic origin group is expressed as a percent of total Hispanics. Because the sample sizes for many groups are too small to make reliable inferences, the differences between the percentages of the specific Hispanic origin groups in the Control Panel compared each of Panels 2-5 were not tested for statistical significance.

4.1.3 What is the effect on the distribution of race and on the reporting of detailed race groups?

Table 7 shows the race distributions for the total population. Race distributions for Hispanics only and for non-Hispanics only are shown in Appendix C. Chi-square tests were used to test for statistical differences between the race distributions in the Control Panel and those for in each of Panels 2-5. Those tests showed that the race distributions in the Control Panel are statistically different from the corresponding distributions in each of Panels 2-5. For the overall population, percentages for individual races in the Control Panel were compared to the corresponding percentages for each of Panels 2-5. Those results are shown in Table 7 and are discussed below. Similar comparisons were not made among the percentages for Hispanics only and for non-Hispanics only due to limited sample size.

Table 8, below, shows the percent among each of four race groups (AIAN, NHPI, SOR, and Asian) who reported a detailed race group. A respondent in the Control Panel could report a detailed race group in one of two ways. They could mark one or more of the check boxes in the extensive list of detailed race groups (e.g., Chinese, Korean, Native Hawaiian) or they could report a detailed race group in the write-in spaces provided. Respondents in Panels 2-5 needed to answer the separate write-in question about ancestry to report a detailed race group.

The results in Tables 7 through 10 show the following:

- NHPI - There were slightly higher percentages of NHPI respondents in Panels 3-5 than in the Control Panel among the overall population. NHPI respondents in Panels 3-5 reported a detailed race group (e.g., Guamanian, Samoan, Fijian) less often than those in the Control Panel.
- Asians - No statistical differences were found in the percentages of Asians among the overall population, but Asian respondents in Panels 2-5 reported a detailed race group (e.g., Asian Indian, Filipino, Japanese) less often than those in the Control Panel. Table 8 lists several detailed Asian groups and shows the percent of Asians who reported each of the group.
- SOR - There were higher percentages of SOR respondents in Panels 2-5 among the overall population and they reported a detailed race group more often in Panels 2, 3 and 5 than in the Control Panel.

Table 7
RACE DISTRIBUTIONS FOR OVERALL POPULATION
(Control Panel & Panels 2-5)
Percent of All Respondents by Panel

Race	Control	2	Panel 3	4	5
White	79.09 (0.34)	75.40* (0.36)	76.79* (.36)	77.19* (0.34)	76.88* (0.37)
Black	8.63 (0.22)	8.78 (0.23)	8.74 (0.24)	8.72 (0.21)	8.41 (0.22)
Asian	4.21 (0.18)	4.08 (0.18)	3.93 (0.18)	4.08 (0.19)	4.25 (0.20)
AIAN	0.53 (0.06)	0.46 (0.05)	0.45 (0.05)	0.48 (0.06)	0.45 (0.05)
NHPI	0.15 (0.04)	0.24 (0.05)	0.27* (0.05)	0.28* (0.06)	0.25* (0.04)
SOR	1.96 (0.13)	4.74* (0.19)	4.45* (0.18)	3.71* (0.17)	4.22* (0.17)
Two+	1.72 (0.11)	2.22* (0.12)	1.87 (0.11)	1.91 (0.12)	2.00* (0.11)
NonResponse	3.72 (0.17)	4.08 (0.16)	3.50 (0.16)	3.63 (0.17)	3.55 (0.15)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.*

Table 8
PERCENT OF RESPONDENTS WITHIN MAJOR RACE GROUPS
WHO REPORTED A DETAILED RACE GROUP
(Control Panel & Panels 2-5)

Percent Within Each Race Group by Panel

Race	Panel				
	Control	2	3	4	5
American Indian or Alaska Native	71.45 (4.92)	63.10 (4.77)	72.92 (4.81)	68.45 (5.24)	79.89 (3.77)
Native Hawaiian or Other Pacific Islander	96.28 (2.83)	58.19* (9.74)	37.53* (7.71)	30.01* (7.27)	31.19* (8.67)
Some Other Race	82.39 (2.67)	87.40* (1.29)	88.72* (1.16)	86.26 (1.36)	90.87* (0.98)
Asian	92.79 (1.24)	84.08* (1.63)	83.56* (1.87)	82.64* (1.77)	83.47* (1.67)
<i>Detailed Asian Race Groups</i> ³					
Asian Indian	15.73 (1.72)	7.35 (1.38)	4.63 (0.95)	7.90 (1.44)	4.05 (0.90)
Chinese	21.03 (1.77)	28.50 (2.08)	25.98 (1.91)	21.67 (1.87)	25.13 (1.89)
Filipino	23.00 (1.89)	13.32 (1.43)	16.60 (1.74)	17.62 (1.72)	18.81 (1.89)
Japanese	10.81 (1.19)	10.08 (1.28)	11.60 (1.31)	10.14 (1.12)	8.03 (1.05)
Korean	9.51 (1.29)	9.10 (1.28)	8.60 (1.29)	9.15 (1.27)	8.18 (1.26)
Vietnamese	6.69 (1.09)	8.98 (1.29)	8.26 (1.27)	9.10 (1.27)	11.55 (1.50)
Other Asian	6.02 (1.17)	6.75 (1.08)	7.89 (1.35)	7.06 (1.32)	7.72 (1.16)

* The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.

³ The percentages for the detailed race groups in this section of the table sum to the total percentages for the total Asian race group. The percent for each detailed Asian race group is expressed as a percent of total Asian respondents. Because the sample sizes of many Asian race groups are too small to make reliable inferences, the differences between the percentages for the Asian race groups in the Control Panel compared with each of Panels 2-5 were not tested for statistical significance.

4.1.4 Summary of Results - Analysis of Three-Question Format

The three-question format (Panels 2-5) resulted in some loss of completeness of data about Hispanic origin.

The information on Hispanic origin is much less complete in the three-question format (Panels 2-5). In the Control Panel, 93.52 percent of Hispanic respondents reported a specific Hispanic origin group. But in Panels 2-5, the percentages of Hispanics who reported a specific Hispanic origin group are at least thirty percentage points lower than in the Control Panel. Examples of specific Hispanic origin groups are Mexican, Cuban, and Puerto Rican.

The percent of Hispanics is slightly higher in the three-question format: 9.02 percent in the Control Panel and between 10.00 percent and 10.35 percent in Panels 2-5. But as mentioned above, Hispanics provided less detailed data about Hispanic origin in Panels 2-5 than in the Control Panel.

The results on the level of response to the Hispanic origin question are mixed. While response to the Hispanic origin question is high on all panels, ranging from 96.39 percent to 97.41 percent, the Hispanic origin response rates for Panels 3 and 4 are higher than in the Control Panel by an estimated 0.76 and 0.64 percentage points, respectively. The response rates of Panels 2 and 5 are not statistically different from the response rates of the Control Panel.

The three-question format (Panels 2-5) produced a higher response to the race question for some respondents but the race distributions differ from the Control Panel.

For the overall population, the response rate to the race question is high in all panels and no differences in the response rate to the race question were found between Panels 2-5 and the Control Panel. The response rates to the race question for all respondents ranged from 95.92 percent to 96.50 percent across the five panels (the Control Panel and Panels 2-5). For Hispanics, the response to the race question differed between the Control Panel and three of the alternative panels (Panels 3, 4 and 5). The race question response rates of Hispanics in these three panels were less than five percentage points higher than the Control Panel. The race question response rate of Hispanics was 69.48 percent in the Control Panel and ranged from 71.92 percent to 74.39 percent across Panels 2-5.

The interpretation of these results is complicated by the fact there were different race distributions in Panels 2-5 compared with those from the Control Panel. There appears to be a shift from reporting race as 'White' to reporting race as 'Some Other Race' in Panels 2-5. Fewer persons reported their race as 'White' than do those from the Control Panel, while about twice as many respondents reported their race as 'Some Other Race' in Panels 2-5. The sum of the percentages of respondents who reported their race as 'White' and 'Some Other Race' in Panels 2-5, are not very dissimilar from the corresponding sum of percentages in the Control Panel. The sum of the percentages in Panels 2-5 ranged from 80.90 percent to 81.24 percent compared

with 81.05 percent for the Control Panel. The magnitude of the shift from reporting race as ‘White’ to reporting it as ‘Some Other Race’ does not appear to be explained by a shift in percentages among the balance of other races groups.

Improving data for use in the Census Bureau’s Intercensal Population Estimates Program was an objective of introducing the three-question format which has a separate question on ancestry. The Intercensal Population Estimates Program and other federal statistical programs that rely on that data do not include an SOR race group category. Census Bureau must impute one of five OMB recognized races to SOR respondents in order to meet the requirements of those programs. The race data collected on the three-question format cannot be used as effectively by the Intercensal Population Estimates Program compared with the race data from the traditional two-question format. This is evident from the percentages of SOR respondents on the Control Panel (1.96 percent) compared with Panels 2-5 (3.71 percent to 4.74 percent). A higher percent of respondents reported race as SOR in each of Panels 2-5 compared with the Control Panel. The percentages of Hispanic SOR respondents are particularly high on Panels 2-5 (26.73 percent to 36.51 percent).

The three-question format (Panels 2-5) resulted in a loss in the completeness of data for Asians and NHPI.

Compared with the Control Panel, the percentages of Asians and NHPI who reported a detailed race group are lower in Panels 2-5. In the Control Panel, 92.79 percent of Asians reported a detailed, while the percentages for Panels 2-5 ranged from 82.64 percent to 84.08 percent. The differences between the Control Panel and each of Panels 2-5 are particularly large for NHPI respondents. In the Control Panel, 96.28 percent of NHPI respondents reported a detailed race group (e.g., Guamanian, Samoan). In Panels 2-5, the percent of NHPI who reported a detailed race group ranged from 30.01 percent to 58.19 percent.

4.2 Analysis of Design Effects on the Traditional Two-Question Format

This section examines the effects of alternative question wording for the Hispanic origin and race questions by contrasting the results from the Control Panel to those from Panel 6. Both panels used the traditional two-question format for capturing race and ethnicity data, showing one question on Hispanic origin followed by one question on race. The differences between two panels are:

- The wording of the initial respondent instruction for responding to the Hispanic origin and races questions differ between the two panels. On the Control Panel it is “Please answer BOTH Questions 8 and 9”. On Panel 6 the instruction is “Please answer BOTH Question 8 about Hispanic origin and Question 9 about race”.

- A note to the respondent that says “Hispanic origins are not races” follows the initial instruction to answer both questions on Panel 6, but the note is not included in the Control Panel.
- The Hispanic origin identifiers were listed in different orders in the Hispanic origin questions. The Control Panel lists them as Spanish, Hispanic or Latino origin, while Panel 6 lists them as Hispanic, Latino, or Spanish origin.
- The Hispanic origin question of Control Panel includes the instruction: ‘Mark “No” if not of Spanish, Hispanic or Latino origin’. This instruction is not included in the Panel 6.
- The race questions of these two panels included slightly different instructions. Panel 6 uses the simple instruction ‘Mark one or more boxes’. The instruction in the Control Panel is ‘Mark one or more races to indicate what this person considers himself/herself to be’.
- The check box responses for the Hispanic origin question were double banked in the Control Panel but single banked in Panel 6.

This analysis does not attempt to explain the effects of each of these individual differences. Combinations of these individual differences within and across the Hispanic origin and race questions may have a confounding effect on the results. Because Hispanic origin and race are closely related topics, effects on the changes to the Hispanic origin question may influence the results on race data and vice versa. The test was not designed to identify the degree to which each wording difference contributed to differences in the results. The results are best interpreted as illuminating the effects of the collective set of all wording differences.

4.2.1 What is the effect on item response rates?

Response to the Hispanic origin question

The wording of Panel 6 decreased the response rate to the Hispanic origin question. The observed response to the Hispanic origin question was 1.65 percentage points lower for Panel 6 than for the Control Panel. This difference is statistically significant at the 10 percent significance level when the null hypothesis is no difference between the two panels.

Table 9
HISPANIC ORIGIN QUESTION
RESPONSE RATES
(Control Panel & Panel 6)
Percent of All Respondents

Panel	Response
Control Panel	96.65 (0.14)
Panel 6	95.00* (0.19)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.*

Response to the race question

Table 10 shows the response to race for the general population, for Hispanics only and for non-Hispanics. No statistically significant differences in the response rates between the two panels were found. This holds true for the general population and for the two sub-populations.

Table 10
RACE QUESTION RESPONSE RATES
(Control Panel & Panel 6)
Percent of All Respondents by Hispanic Origin

	All Respondents	Hispanics	Non-Hispanics
Control Panel	96.28 (0.17)	69.48 (1.36)	99.40 (0.05)
Panel 6	95.92 (0.17)	71.19 (1.29)	99.31 (0.07)

4.2.2 What is the effect on the distribution of Hispanic respondents and on the reporting of specific Hispanic groups?

Table 11 shows the percent of the population that reported they were Hispanic and the percent of the population that provided a specific Hispanic origin in the Hispanic origin question. Respondents could provide a specific Hispanic origin in two ways. The first is checking at least

one of the three detailed origin response boxes: 1) Mexican/ Mexican American/Chicano, 2) Puerto Rican, or 3) Cuban. The second is checking the “Yes, another Hispanic, Latino, or Spanish origin” response and writing in a specific Hispanic origin. The coded write-in responses were tabulated for respondents who mark this box. Respondents who checked this box but did not provide a write-in response that could be coded as a Hispanic origin were treated as Hispanic. However these persons were not treated as having provided a specific Hispanic origin.

Table 12 shows the percent of Hispanics who provided a specific Hispanic origin and the distribution of reported Hispanic origins based on the tabulation of check box responses and coded write-in responses. The first row of this table shows the percent of Hispanics who did not report a specific Hispanic origin as a write-in response.

A higher percent of the population responded that they were Hispanic in Panel 6. Also, the results in Tables 13 and 14 show that data on Hispanic origin were more complete in Panel 6 than in the Control Panel. This can be seen from the differences in the percentages of the total population who provided a specific Hispanic origin (Table 13) and differences in the percentages of Hispanics who provided a specific Hispanic origin (Table 14).

Table 11
PERCENT OF RESPONDENTS THAT ARE HISPANIC⁴
(Control Panel & Panel 6)
Percent of All Respondents

	Control Panel	Panel 6
Hispanics	9.02 (0.26)	9.72* (0.27)
Hispanics Who Reported a Specific Hispanic Origin	8.44 (0.25)	9.34* (0.26)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level.*

⁴ Note that the percent of Hispanics in the population as shown in Table 5 is an estimate of the distribution of Hispanics in mail response households. Based on Census 2000 data, the expected percent for the entire U.S. population will be larger.

Table 12
PERCENT OF HISPANIC RESPONDENTS WHO REPORTED
A SPECIFIC HISPANIC ORIGIN
(Control Panel & Panel 6)
Percent of Hispanics

	Control Panel	Panel 6
No Specific Origin Reported	6.48 (0.65)	3.93* (0.49)
Total Hispanics Who Reported a Specific Hispanic Origin	93.52 (0.65)	96.07* (0.49)
<i>Distribution of Specific Origin Groups⁵</i>		
Mexican	59.25 (1.15)	62.09 (1.34)
Puerto Rican	11.70 (0.89)	10.83 (0.80)
Cuban	4.35 (0.49)	5.32 (0.80)
Dominican	1.34 (0.31)	2.39 (0.41)
Central American	5.54 (0.61)	6.23 (0.65)
South American	7.07 (0.78)	6.07 (0.64)
Latin American	0.02 (0.02)	0.09 (0.06)
Other Hispanic group	4.26 (0.50)	3.05 (0.40)

* The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level

⁵ The percentages in this section of the table sum to the total percentages for “People Who Reported a Specific Origin”. The percent for each specific Hispanic origin group is expressed as a percent of total Hispanics. Because the sample size for many groups are too small to make reliable inferences, the differences between the percentages of the specific Hispanic origin groups in the Control Panel compared with Panels 6 were not tested for statistical significance.

4.2.3 What is the effect on the distribution of race and on the reporting of a detailed race group?

Table 13 shows the distribution of race responses for the total population, for Hispanics only and for non-Hispanics only. Chi-square tests were used to test for differences between the race distributions in the Control Panel and those in Panel 6. This test showed that each race distribution in the Control Panel is statistically different from the corresponding distribution in Panel 6. For the overall population, percentages for individual races in the Control Panel were compared to the corresponding percentages in Panel 6. Those results are shown in Table 13 under the column for ‘All Respondents’. Similar comparisons were not made among the percentages for Hispanics only and for non-Hispanics only due to the limited sample size.

Table 13⁶
RACE DISTRIBUTIONS
(Control Panel & Panel 6)

Race	Percent of Respondents by Hispanic Origin					
	All Respondents		Hispanics		Non-Hispanics	
	Control Panel	Panel 6	Control Panel	Panel 6	Control Panel	Panel 6
White	79.09 (0.34)	79.07 (0.37)	46.09 (1.47)	53.01 (1.38)	83.56 (0.34)	83.80 (0.35)
Black	8.63 (0.23)	8.69 (0.24)	2.32 (0.40)	1.72 (0.25)	8.80 (0.24)	8.59 (0.26)
Asian	4.21 (0.18)	4.52 (0.19)	0.26 (0.11)	0.75 (0.23)	4.49 (0.20)	4.70 (0.20)
American Native or Alaska Native	0.53 (0.06)	0.33* (0.05)	1.35 (0.29)	1.07 (0.25)	0.45 (0.06)	0.23 (0.04)
Native Hawaiian or Other Pacific Islander	0.15 (0.04)	0.08* (0.02)	0.12 (0.08)	0.06 (0.06)	0.14 (0.04)	0.08 (0.02)
Some Other Race	1.96 (0.13)	1.63* (0.12)	15.78 (1.07)	11.92 (0.95)	0.48 (0.07)	0.45 (0.06)
Two or More Races	1.72 (0.11)	1.61 (0.11)	3.56 (0.57)	2.68 (0.36)	1.49 (0.10)	1.45 (0.12)
NonResponse	3.72 (0.17)	4.08 (0.17)	30.52 (1.36)	28.81 (1.29)	0.60 (0.05)	0.69 (0.07)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level*

⁶ Because the sample size of many race groups within the ‘Hispanics’ and ‘Non-Hispanics’ portions of the table are too small to make reliable inferences, the differences between the percentages of individual race groups in the Control Panel and each of Panels 2-5 were not tested for statistical significance.

Table 14 shows, for each of four races, the percent who reported a detailed race group. There is little evidence of a difference in completeness of data for the two questionnaires. The NHPI race group is the only race group for which there is a statistical difference in the percent who report a detailed race group. Respondents could report a detailed race group by either marking one of the race group check boxes in the race question or by writing a detailed race group in the spaces that accompany the AIAN, NHPI, Other Asian, and SOR check boxes. A written detailed race group responses was tabulated only if it corresponded with the race of the check box marked.

Table 14
PERCENT OF Respondents WITHIN MAJOR RACE GROUPS
WHO REPORTED A DETAILED RACE GROUP
(Control Panel & Panel 6)
Percent for Each Race Group by Panel

Race	Control Panel	Panel 6
American Indian or Alaska Native	71.45 (4.92)	61.65 (6.75)
Native Hawaiian or Other Pacific Islander	96.28 (2.83)	73.00* (10.09)
Some Other Race	82.39 (2.67)	86.85 (2.28)
<hr style="border-top: 1px dashed black;"/>		
Asian	92.79 (1.24)	94.05 (1.31)
<i>Detailed Asian Race Groups</i> ⁷		
Asian Indian	15.73 (1.72)	15.91 (1.81)
Chinese	21.03 (1.77)	24.33 (1.83)
Filipino	23.00 (1.89)	20.46 (1.78)
Japanese	10.81 (1.19)	9.81 (1.13)
Korean	9.51 (1.29)	10.54 (1.39)
Vietnamese	6.69 (1.09)	8.11 (1.30)
Other Asian	6.02 (1.17)	4.89 (0.94)

** The difference between this value and the value for the Control Panel is statistically significant at the 10 percent significance level*

⁷ The percentages for the detailed race groups in this section of the table sum to the total percentages the total Asian race group. The percent for each detailed Asian race group is expressed as a percent of total Asian respondents. Because the sample size for many groups are too small to make reliable inferences, the differences between the percentages of the detailed Asian race groups in the Control Panel and each of Panels 2-5 were not tested for statistical significance.

4.2.4 Summary of Results - Analysis of Design Effects on Traditional Two Question Format

Data on Hispanic origin are more complete in Panel 6.

The response to the Hispanic origin question was high on both the Control Panel (96.65) and Panel 6 (95.00). A higher percent of respondents reported they are Hispanic in Panel 6 (9.72 percent) than in the Control Panel (9.02 percent). Because the percent of Hispanics was higher in Panel 6 but the response to the Hispanic question was lower in Panel 6, we know that non-Hispanics skipped the Hispanic origin question in Panel 6 relatively more often than Hispanics did. This shows that non-Hispanics contributed more to the lower response in Panel 6 than did Hispanics.

More complete Hispanic origin data were collected in Panel 6 compared with the Control Panel. The percent of Hispanics who reported a specific Hispanic origin was 96.07 percent in Panel 6 and 93.52 percent in the Control Panel.

It is difficult to know what features of the two versions of the Hispanic origin question influenced these results, since there were multiple differences between the two questionnaires with respect to the wording and the format. The phrase “Mark ‘No’ if not of Spanish, Hispanic or Latino origin” was included in the Control Panel but not in Panel 6, and the order in which the Hispanic identifiers (Spanish, Hispanic, and Latino) were listed differs. Also, the Hispanic origin question check box response categories are double banked in the Control Panel but single banked in Panel 6. It is not possible to determine how much each of these differences contributed to the results independently or in combination. However, the results clearly point to more complete Hispanic origin data for respondents in Panel 6.

Both the Control Panel and Panel 6 perform equally well with respect to race data.

No statistical differences were found in the response rates to the race question between the Control Panel and Panel 6. Also, there was little difference in the completeness of the race data. With the exception of the NHPI race group, there were no statistical differences between the panels in the percent who reported a detailed AIAN, Asian, or SOR race group. NHPI respondents reported a detailed race group 96.28 percent of the time in the Control Panel and 73.00 percent of the time in Panel 6.

4.3 Analysis of Question Design Effects Within the Three-Question Format

This section analyzes the effects of differences between the questions of the three-question format. Four design effects are analyzed. These are:

- The inclusion of examples in the Hispanic origin and race questions
- The inclusion of an instruction to emphasize that Hispanic origins are not races
- A separate question on American Indian or Alaska Native ancestry or tribe
- A modified ancestry question that includes a sentence to clarify what is meant by ancestry.

4.3.1 What is the effect of including examples in the Hispanic origin and race questions?

In this section we compare results for Panel 2 with those for Panel 3 to examine if including examples in the Hispanic origin and race questions promotes higher response rates and more complete reporting of race and ethnicity. The questionnaires for Panels 2 and 3 were identical except that examples of Hispanic origin and examples of races were included in the Hispanic origin and race questions of Panel 2 but not of Panel 3.

We examined five metrics to analyze the effects of adding examples to the questions. We examined the response rates for the Hispanic origin and race questions, the frequency of reporting specific Hispanic origins among Hispanics, the frequency of reporting detail on races, and the percent who reported they are Hispanic.

Including the examples slightly reduces the response rates to the Hispanic origin and race questions. While the response rate to these two questions was high on both panels, there was a slightly higher response to both the Hispanic origin and race questions on the questionnaire that did not include examples in either question. No differences were found in any of the other metrics.

Response to the Hispanic Origin Question

The response rates to the Hispanic origin question differed between the two questionnaires. The response rates were 96.83 percent and 97.41 percent for Panels 2 and 3, respectively (see Table 3). These rates were found to be statistically different at the 10 percent level of significance.

Response to the Race Question

The response rates to the race question differed between the two questionnaires. The response rates were 95.92 percent and 96.50 percent for Panels 2 and 3, respectively (see Table 4). These percentages were found to be statistically different at the 10 percent level of significance.

Reporting Specific Hispanic Origin

We examined the percent of Hispanics who reported a specific Hispanic origin and found no differences between the percentages on these two questionnaires. The percentages of Hispanics who reported a specific Hispanic origin were low on both questionnaires, 59.81 percent and 62.25 percent in Panels 2 and 3, respectively (see Table 6) .

Reporting of Detailed Race Groups

Among those who reported their race as AIAN, Asian, NHPI, or SOR, we examined the percent of times these respondents reported a detailed race group. We found no differences between the questionnaires in these percentages. The percent who reported a detailed race group were 81.38 (s.e. 0.97) percent and 82.44 (s.e. 0.99) percent for Panels 2 and 3, respectively.

Percent of Hispanics

No difference was found between the two questionnaires in the percentages who reported they are Hispanic. The percentages are 10.24 percent and 10.35 percent for Panels 2 and 3, respectively (see Table 5).

4.3.2 What is the effect of including an instruction to emphasize that Hispanic origins are not races?

In this section, we compare results for Panel 3 with those for Panel 4. The race questions differed only in that the race question in Panel 4 included the additional instruction “For this census, Hispanic origins are not races”. The Hispanic origin questions are identical in these two panels.

We examined four metrics to analyze the effects of adding the instruction to the race question. These are the response rates to the Hispanic origin question, the response rates to the race question for all respondents and for Hispanics only, and the percent who reported they are Hispanic.

The data show that including the instruction had no effect on the results. No differences were found in the rate of response to either the Hispanic origin or race question, nor was there a difference in the percent who reported they are Hispanic.

Response to the Hispanic Origin Question

There was no difference in the response rates for the Hispanic origin question between the two questionnaires. The response rates to Hispanic origin were 97.41 percent and 97.29 percent for Panels 3 and 4, respectively (see Table 3).

Response to the Race Question

The response rates for the race question did not differ between the two questionnaires. The response rates were 96.50 percent and 96.37 percent for Panels 3 and 4, respectively (see Table 4). These percentages were found to be statistically different at the 10 percent level of significance.

Percent of Hispanics

No difference was found between the two questionnaires in the percentages of who reported they are Hispanic. The percentages were 10.35 percent and 10.08 percent for Panels 3 and 4, respectively (see Table 5).

4.3.3 What is the effect on the completeness of data from including a separate ancestry question and a separate question about American Indian or Native American tribe?

In this section we compare the results for Panel 4 to Panel 5. The Hispanic origin and race questions are identical in the two panels. Panel 5 included an additional question that asks AIAN respondents to report their tribe as well as whether or not they are enrolled in a tribe. The ancestry question was modified to reflect the inclusion of the separate question on AIAN tribe. This panel was included to obtain information about whether the ancestry question alone was sufficient to obtain information about tribal affiliation, in the absence of a separate question about tribe.

We examined three metrics to analyze the effects of using a separate ancestry question to capture tribal information for AIAN respondents. We examined the percentages of who reported they are AIAN, the percent of AIAN respondents who reported a tribe and the percentages of Asian, NHPI and SOR respondents who reported a detail race group.

No difference was observed in the percent of AIAN respondents. While the percent of AIAN respondents who reported a tribe is low on both panels, the percent for Panel 5 was statistically higher than it was for Panel 4.

Percent of AIAN Respondents

No difference was found between the two questionnaires in the percentages of who reported they are American Indian or Alaska Native. The percentages are 0.48 percent and 0.45 percent for Panels 4 and 5, respectively (see Table 7).

Percent of AIAN Respondents Who Report Tribe

The use of a separate ancestry question to collect information on tribe increased the percent of AIAN respondents who report a tribe. But, the values for both Panels 4 and 5 were somewhat low, 68.45 percent and 79.89 percent, respectively (see Table 8) .

Reporting of Specific Race Subgroups

We examined the percent of times respondents who reported their race as Asian, NHPI and SOR, reported a detailed race group. We found no differences between the questionnaires for these respondents. The percentages of who reported a detailed race group were 79.55 (s.e. 1.11) percent and 81.22 (s.e. 1.00) percent for Panels 4 and 5, respectively.

5. REFERENCES

Angueira, Teresa (June 9, 2004), 2010 Census Planning Memorandum Series: Action Plan: 2010 Research and Development Planning Group on Race and Ethnicity Data Collection, Tabulation and Editing

Bennett, Claudette (February 7, 2005), Revised Final Wording for the Race and Ethnicity Questions for the 2005 National Census Test

Bouffard, Julie (January 11, 2006), Sample Weights for the 2005 National Census Test - Updated

Killion, Ruth Ann (March 9, 2005), Planning, Research, and Evaluation Division, TXE/2010 Memorandum Series: DCO-NCT2005-S-01, Sample Specifications for the 2005 National Census Test

Martin, Elizabeth; Sheppard, David; Bentley, Michael; Bennett, Claudette (September 7, 2004) , Results of the 2003 National Census Test of Race and Hispanic Questions

Sheppard, David; Bentley, Michael; Woltman Henry (March 3, 2004), 2003 National Census Test: Analysis of Alternative Race and Ethnicity Questions Based on Panel Comparisons

Supporting Statement U.S. Census Bureau, 2005 National Census Test (Submitted to the Office of Management and Budget)

Tancreto, Jennifer (January 25, 2006), DSSD 2005 Memorandum Series #E-2, 2005 National Census Test: Global Test Measures and Issues During Production

2005 National Census Test (NCT), Race and Ethnic Test Objectives (June 9, 2005, Baseline Version 1.0)

2005 National Census Test, Panel Design Proposal (February 17, 2005)

2005 National Census Test Project Management Plan (Version 6.0, July 6, 2005)

Summary of Alternatives and Treatments for Hispanic Origin and Race

TREATMENTS	CONTROL	ALTERNATIVES				
	1	2	3	4	5	6
Two-Question Format						
Census 2000-style question	X					X
Note to respondents included an instruction to answer both the Hispanic origin and race questions only	X					
Note to respondents included an instruction to answer both the Hispanic origin and race questions and a statement saying that – <i>For this census, Hispanic origins are not races.</i>						X
Reordered Hispanic origin identifiers						X
Single-banked Hispanic origin response categories						X
Included instruction – <i>Mark [X] “No” if not of Spanish, Hispanic, or Latino origin.</i>	X					
Different race instructions – For Control: <i>Mark [X] one or more races to indicate what this person considers himself/herself to be.</i> For Alternative 6: <i>Mark [X] one or more boxes.</i>	X					X
Examples in Hispanic origin and race questions	X					X
Three-Question Format						
Shortened Hispanic origin and race questions combined with added ancestry question		X	X	X	X	
Examples in Hispanic origin and race questions		X				
Included a modified race question instruction - <i>“For this census, Hispanic origins are not races.”</i>				X		
Note to respondents included an instruction to answer both the Hispanic origin and race questions only				X		
Included a modified ancestry question with a narrative lead in				X	X	
Separate question for American Indian or Alaska Native tribe					X	

Table 15⁸
RACE DISTRIBUTIONS FOR HISPANICS
(Control Panel & Panels 2-5)
Percent of Hispanics

Race	Control	2	Panel 3	4	5
White	46.09 (1.47)	26.72 (1.12)	33.91 (1.29)	39.40 (1.32)	35.47 (1.21)
Black	2.32 (0.40)	2.04 (0.30)	1.88 (0.31)	1.87 (0.29)	1.70 (0.29)
Asian	0.26 (0.11)	0.86 (.023)	.54 (0.13)	0.74 (0.24)	0.38 (0.12)
AIAN	1.35 (0.29)	0.70 (0.18)	0.89 (0.23)	0.88 (0.30)	0.87 (0.20)
NHPI	0.12 (0.08)	0.31 (0.12)	0.09 (0.04)	0.37 (0.15)	0.22 (0.1)
SOR	15.78 (1.07)	36.51 (1.26)	32.81 (1.28)	26.73 (1.26)	31.66 (1.24)
Two+	3.56 (0.57)	4.78 (0.55)	3.51 (0.45)	3.37 (0.45)	4.07 (0.47)
NonResponse	30.52 (1.36)	28.08 (1.15)	26.38 (1.19)	26.65 (1.24)	25.61 (1.12)

⁸ Because the sample sizes for many categories are too small to make reliable inferences, the differences between the percentages in the Control Panel and each of Panels 2-5 were not tested for statistical significance.

Table 16⁹
RACE DISTRIBUTION FOR NON-HISPANICS
(Control Panel & Panels 2-5)
Percent of Non-Hispanics

Race	Control	2	Panel 3	4	5
White	83.56 (0.34)	81.92 (0.33)	82.46 (0.34)	82.35 (0.33)	82.41 (0.35)
Black	8.80 (0.24)	9.37 (0.25)	9.29 (0.26)	9.22 (0.23)	8.85 (0.24)
Asian	4.49 (0.20)	4.46 (0.20)	4.29 (0.19)	4.41 (0.21)	4.65 (0.21)
AIAN	0.45 (0.06)	0.41 (0.05)	0.39 (0.05)	0.40 (0.05)	0.38 (0.05)
NHPI	0.14 (0.04)	0.24 (0.05)	0.29 (0.05)	0.37 (0.06)	0.25 (0.05)
SOR	0.48 (0.07)	1.03 (0.09)	1.11 (0.10)	1.05 (0.10)	1.02 (0.09)
Two+	1.49 (0.10)	1.86 (0.11)	1.65 (0.11)	1.73 (0.12)	1.68 (0.11)
NonResponse	0.60 (0.05)	0.72 (0.06)	0.53 (0.05)	0.56 (0.06)	0.75 (0.06)

⁹ Because the sample sizes for many categories are too small to make reliable inferences, the differences between the percentages in the Control Panel and each of Panels 2-5 were not tested for statistical significance.

2005 National Census Test Overview

In preparation for the 2010 Census, the Census Bureau is conducting a series of tests. In late 2005, a mailout/mailback national test was conducted using variations of questionnaire content, and various methods to increase response to the Census, including replacement questionnaire methods. The test also included the Internet as an optional mode for completing the census short form. Census Day was September 15, 2005.

The objectives for the 2005 National Census Test were:

- Test methods to improve completeness and accuracy of reporting for short form items, including tenure, relationship, age and date of birth, and race and Hispanic origin.
- Test ways to reduce respondent and data capture errors, and improve respondent friendliness in mail and Internet modes.
- Test ways to improve coverage accuracy by reducing omissions and erroneous enumerations, and/or flagging potential errors for coverage followup interviews.
- Test ways to improve the operational feasibility of the second mailing.
- Test ways to improve self-response and maintain data quality by mailing bilingual questionnaires.