2005 National Census Test:
Results for "Space Saving" Format Changes
And Reversal of Age and Date of Birth Items

Elizabeth A. Martin
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: Results for "Space Saving" Format Changes And Reversal of Age and Date of Birth Items

Elizabeth A. Martin
Director's Office

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

CONTENTS

List of Tables

EXECUTIVE SUMMARY

The 2005 National Census Test (NCT) was a mailout/mailback experiment conducted to evaluate alternative questions and formats intended to improve the accuracy of census count and characteristic data, and alternative mailing and contact strategies intended to improve self-response rates.

The sample was stratified into areas with high and low concentrations of non-Whites or Hispanics. High concentration areas were oversampled.

This report evaluates several format and wording changes that were intended to free up space on the form, or to simplify question wordings and response formats. Experimental changes that result in data of *equal quality* can be implemented if it is desirable to do so for other reasons (e.g., to free up space on the form). Any changes that result in data of *better quality* are recommended for adoption.

This report also evaluates the effects of reversing age and date of birth items. In automated instruments, it is advantageous to ask date of birth first and then confirm age, but the mail questionnaire has traditionally asked age first. The NCT included an experiment to evaluate the two orders, in order to determine whether this inconsistency between instruments in different modes is acceptable. If the age-first sequence obtains *better data or data of the same quality* as the age-second sequence, then the current, preferred order (age first) is recommended for retention in the mail questionnaire. If the age-first sequence obtains *worse data*, then the reversed order is recommended for adoption in the 2010 census.

Name

Two experiments in the NCT involve the request for names of household members:

- Abbreviated formats for recording names were evaluated, to save space on the form
- Spelling out "MI" (middle initial) for Person 1 was tested, to obtain more complete name information and facilitate matching and unduplication.

Results show that:

- For Person 1, the abbreviated name format increased missing data and is not recommended.
- For Persons 2-6, the abbreviated format resulted in a slight loss of name information in the high non-White or Hispanic stratum, but did not affect the completeness of name reporting for the sample as a whole. This change is deemed acceptable if desirable on other grounds (i.e., to save space on the form).
- Spelling out middle initial for Person 1 resulted in much more complete reporting of middle initials. There was no carryover effect on reporting of middle initials for Persons 2-6.

Sex

The NCT included an experiment that eliminated the instruction to "Mark ONE box" and placed the response categories to the right of the instruction, to save space on the form.

Results show that the change led to a substantial loss of information about sex for Persons 2-6. It is not recommended.

Age—Format

The NCT included an experiment that shortened and simplified the age question, and reformatted the answer spaces and labels.

Results show that the changes led to more complete information about date of birth, but deterioration in reported age. The change is not recommended.

Age—Order

The NCT included an experiment that reversed the sequence of the age and date of birth items.

The results show that asking date of birth first improves completeness of date of birth reporting, with no effects on reported age, age distributions, or consistency of date of birth and reported age. It is recommended that the mail questionnaire ask date of birth first, as is done in other decennial automated instruments.

Hispanic Origin

The Hispanic Origin item was simplified by eliminating an instruction ("Mark 'no' if not of Spanish, Hispanic, or Latino origin") and single- rather than double-banking response categories. These changes were tested in the presence of an enhanced instruction to respondents to provide answers to both Hispanic origin and race questions (see Alberti, 2006 for the full analysis of the experimental changes to the race and Hispanic origin questions).

Results show that these changes led more people to check a box identifying themselves as Hispanic (and specifically as Mexican). More non-Hispanics skipped the item, resulting in an increase in overall item nonresponse. It is recommended that an experiment be conducted in 2010 to isolate the effects of instructional, formatting, and wording variations on Hispanic reporting.

1. INTRODUCTION

a. 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 (NCT) 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.

This analysis addresses the first two objectives.

b. Objectives of this Analysis

In an effort to save space on the Census form and to simplify the response task, Panel 6 (Form DC-1F) of the NCT incorporates alternative formats and wordings of questions asking for name, sex, race, Hispanic origin, date of birth (DOB) and age. In addition, "Middle initial" is spelled out for Person 1, for the purpose of increasing the availability of this information for matching and unduplication. These changes (described in more detail below) are tested in the presence of a revised respondent note to answer questions on both race and ethnicity.

This analysis examines whether the changes affect data completeness and quality, compared to the control. This report compares the so-called "space saving" panel (panel 6, form DC-1F) against the control (id structure 1, form DC-1A). For tenure, panel 2 (form DC-1B) is used as the control. (Note that not all of the experimental changes in panel 6 did or were intended to save space; some were meant to simplify question wordings or formats or make other changes to reduce respondent burden.)

Experimental changes that result in data of *equal quality* can be implemented if it is desirable to do so for other reasons (e.g., to free up space on the form). Any changes that result in data of *better quality* are recommended for adoption.

In the automated versions of the instruments being tested for the 2010 census, date of birth is asked first, then age. This makes it possible for interviewers to calculate age and confirm it

with respondents, rather than ask two seemingly redundant questions. This sequence departs from the traditional order of asking age then date of birth in the mail form.

It would be desirable to maintain the same item order in all instruments, to avoid introducing mode differences in question context. Past research showed, however, that reversing the traditional sequence results in slightly higher item nonresponse rates for age and date of birth (Spencer and Perkins, 1998). Panel 4 (form DC-1D) in the 2005 NCT reverses the sequence of items to evaluate the effects on reporting.

If the age-first sequence obtains *better data or data of the same quality* (i.e., the same or lower item nonresponse rates and rates of inconsistency) as the age-second sequence, then the current, preferred order (age first) will be recommended for retention in the mail questionnaire.

If the age-first sequence obtains *worse data* (i.e., higher item nonresponse rates, higher inconsistency) than the age-second sequence, then the reversed order will be recommended for adoption in the 2010 census. In this case, additional analysis is needed to assess distributional effects of the order reversal.

2. METHODOLOGY

The sample for the 2005 NCT was selected from housing units that resided in mailout areas (areas with city-style addresses) on the Hundred-Percent Census Edited File (HCEF). About 80 percent of all housing units in the United States had city-style addresses as of Census 2000. Puerto Rico was not part of the test.

A total of 420,000 housing units were selected across 24 panels. The sample size for each of the four panels analyzed in this report was 30,000 housing units.

The sampling frame was stratified into areas with high and low concentrations of people of Black, Asian, Pacific Islander, or American Indian and Alaska Native race, or Hispanic origin (Bentley, 2005). Areas with high non-White or Hispanic concentrations were oversampled. All results in this report are weighted.

The 2005 NCT used multiple mailings to contact sampled housing units. Every housing unit was sent an advance letter as a first contact around August 22, 2005.

A week later (August 29, 2005) the initial questionnaire package was mailed. It contained a questionnaire, a postage-paid return envelope, and a letter from the Census Bureau's Director that encouraged households to respond and provided the option of responding by Internet.

A reminder postcard was sent on September 6, 2005. It included a statement reminding households to respond to the NCT if they had not already done so, and provided instructions so that households could use the Internet to respond.

Finally, a replacement questionnaire that contained identical content to the initial questionnaire was sent to housing units that had not responded prior to September 13, 2005. Accompanying the questionnaire was a letter from the Director urging response and

providing instructions for using the Internet. The replacement questionnaire package was delivered to the post office for mailing on September 23, 2005.

The initial mailing coincided with Hurricane Katrina, which disrupted mail service to and from affected areas of the country. Replacement questionnaires were not mailed to those areas, and cases in that area were treated as Undeliverable as Addressed (Tancreto 2006).

3. LIMITATIONS

Multiple formatting and wording changes were tested in combination, because it was not feasible to test them separately. This confounding implies that we cannot be certain of the causes of some of the results reported here. In interpreting the findings, I identify probable causes of the results, but experimental confounding limits the certainty that can be placed in any conclusions.

As noted above, areas that were affected by Hurricane Katrina are underrepresented in the results.

The NCT is a test, not the census itself. Response rates are lower for census tests, which lack census publicity.

4. RESULTS

Each topic (name, sex, age format, age order, Hispanic origin, tenure) is discussed in a separate section that describes the experimental treatment, presents results, and discusses recommendations. The analysis is restricted to data defined people on mail returns; most analyses are restricted to Persons 1-6 since the experimental manipulations were only carried out for questions asked about Persons 1-6. (An exception is the age/date of birth order experiment, which was also implemented in the continuation roster for Persons 7-12.)

Mail response rates for the panels analyzed in this report are 53.8 percent for the control and 53.3 percent for each of the three experimental panels[1].

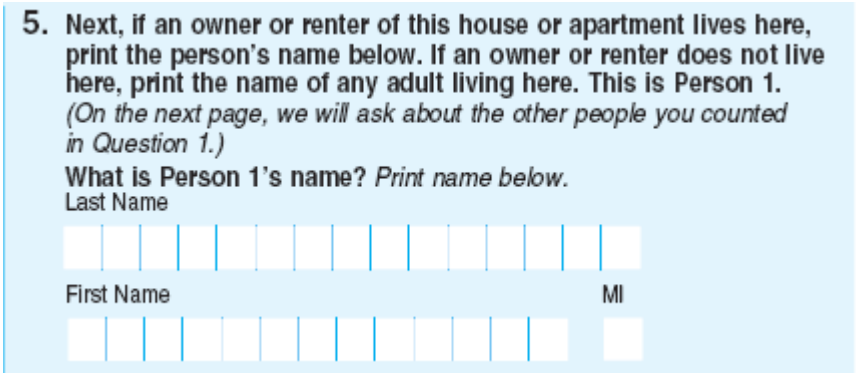
Standard errors are computed using stratified jackknife replication methods in VPLX, and two-tailed t-tests are used to evaluate differences between panels. In the tables, percentages shown in bold represent statistically significant ($p<.10$) differences between experimental and control panels.

[1] The *mail response rate* is a measure of respondent behavior with regard to responding to the census test. The numerator is the number of sample cases for which we received a nonblank mail return. The denominator is the number of sample cases minus the number of sample cases identified by the USPS as "undeliverable as addressed" (UAA). This calculation is comparable to cooperation rates computed in previous census tests, including other evaluations of the 2005 NCT, the 2003 National Census Test, the Census 2000 experiments, and the 1992 and 1993 Census Tests.

a. NAME

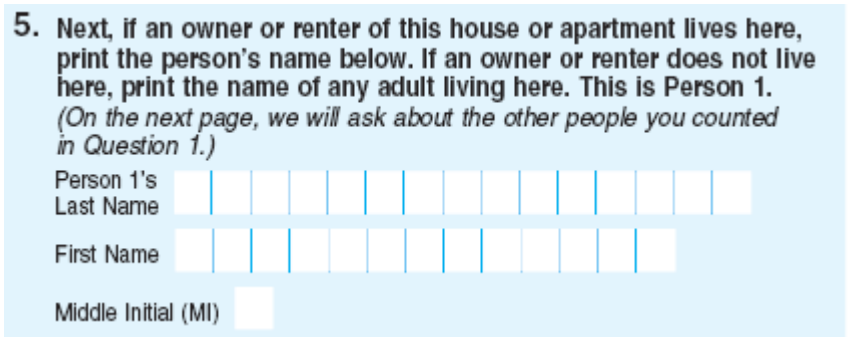
i. Experimental Treatment

Name formats are different for Person 1 and subsequent people listed on the form. The name format in the control form is virtually identical to the format used in Census 2000, although the preceding instruction was modified based on the results of cognitive testing:

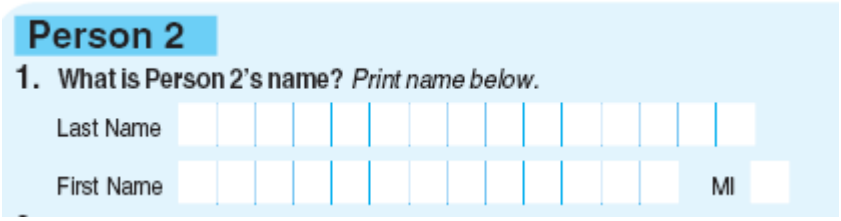

5. **Next, if an owner or renter of this house or apartment lives here, print the person's name below. If an owner or renter does not live here, print the name of any adult living here. This is Person 1.**
(On the next page, we will ask about the other people you counted in Question 1.)
What is Person 1's name? *Print name below.*
Last Name
First Name MI

The request for Person 1's name is shortened in Panel 6, by eliminating the separate question and instruction, ("What is Person 1's/this person's name? *Print name below.*"), since "Print the person's name" is already part of the instruction. The instruction about whom to list as Person 1 is identical to the control, but the write-in spaces are placed adjacent to rather than beneath the labels, and "Middle Initial" is spelled out and moved below the "First Name" write-in space.

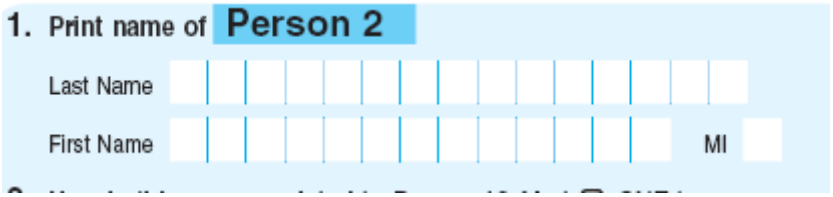
Missing data rates are typically low for first and last names (see, e.g., Table 2), but much higher for middle initials. It was hypothesized that the high missing rate might occur because people either did not notice or did not absorb the intended meaning of "MI", so they left the box blank. The placement of MI to the right, out of the respondent's line of sight, probably contributes to it being overlooked. Spelling out MI the first time it is presented (for Person 1) was intended to train respondents what MI stands for, so they would fill in middle initials for Person 1 and subsequent people.


5. **Next, if an owner or renter of this house or apartment lives here, print the person's name below. If an owner or renter does not live here, print the name of any adult living here. This is Person 1.**
(On the next page, we will ask about the other people you counted in Question 1.)
Person 1's Last Name
First Name
Middle Initial (MI)

The request for names for Persons 2-6 is formatted as follows in the control form:



This is shortened in the experimental form by incorporating "Print name of…" into the label heading each person column, and eliminating the question. Middle initial is not spelled out, due to lack of space on the form.



ii. Results

Table 1 reports results of these manipulations separately for Person 1 and for Persons 2-6.

Table 1. Percentage of Cases Missing Name Information, by Experimental Treatment

	Control	Experimental (6)	t_{E-C}
Person 1			
Last Name	**1.02%** (.0912)	**1.42%** (.1015)	3.03**
First Name	**1.40%** (.1028)	**1.81%** (.1185)	2.66**
Middle Initial	**32.66%** (.4110)	**16.61%** (.3425)	-30.91**
Persons 2-6			
Last name	0.80% (.1008)	0.87% (.1022)	0.54
First name	0.82% (.0934)	0.95% (.1005)	0.91
Middle Initial	31.40% (.4647)	31.04% (.4897)	-0.54

**p<.01

The experimental changes resulted in a significant loss of information about first and last names for Person 1. The lack of an identifiable question following a long instruction evidently was problematic, leading some respondents to skip Person 1's name.

On the other hand, there is essentially no difference between panels in the completeness of name reporting for Persons 2-6. This might suggest that this space-saving modification is acceptable if desirable for other reasons. However, there is a marginally significant loss of last name information in the experimental form in the high non-White or Hispanic stratum, as shown in Table 2.

Table 2. Percentage of Cases Missing Last Name Information for Persons 2-6, by Experimental Treatment and Stratum

Stratum	Control	Experimental	t_{E-C}
High nonwhite or Hispanic	**0.74%** (.1247)	**1.16%** (.1822)	1.89*
Low nonwhite or Hispanic	0.82% (.1284)	0.77% (.1226)	-.26

*p<.10

Spelling out MI and placing it below first name substantially increased the completeness of reporting middle initials for Person 1. As shown in Table 1, the nonresponse rate was cut in half, from about 33 percent to 17 percent. Spelling out MI for Person 1 had no carryover effect for subsequent people. About 31 percent of Persons 2-6 failed to report middle initials

in both treatments. This suggests that MI is overlooked, rather than not understood. If the latter were the problem, we would expect spelling it out for Person 1 to have a carryover effect for Persons 2-6.

Note also that first and last name information is significantly more complete for Persons 2-6 than for Person 1, in both forms.

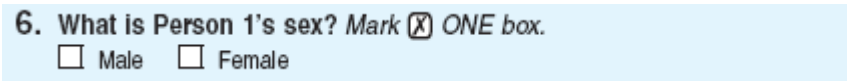
iii. Recommendations

- The formatting changes made for Person 1's name led to loss of data and should not be adopted. In particular, the bolded question, "What is Person 1's name?" should be retained for Person 1.
- Reformatting name as it was done for Persons 2-6 in the experimental form did not result in loss of data for the sample as a whole, but it did lead to some loss of data for cases in the high non-White or Hispanic stratum.
- Spelling out "middle initial" and moving it below first name for Person 1 resulted in much more complete reporting, and would increase the data available for matching. Lynch (2005) notes that middle initial has very high discriminating ability for matching, and should be used in future matching endeavors. (As will be argued below, it may have also led to more complete sex information.) This change is recommended if feasible due to space constraints.

b. SEX

i. Experimental Treatment

In the control form, sex contains an instruction ("*Mark [X] ONE box.*") that was used in Census 2000. This instruction may be a relict of the 1990 census, when sex was the first item on the form that required a circle to be filled, hence requiring an instruction to train respondents how to answer such questions.


6. What is Person 1's sex? *Mark* (X) *ONE box.*
☐ Male ☐ Female

Experimental panel 6 eliminates the instruction, and places the response categories to the right rather than below the question:


6. What is Person 1's sex? ☐ Male ☐ Female

ii. Results

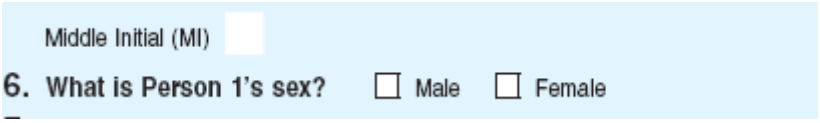
Table 3 shows that the experimental treatment has different effects for Person 1 and subsequent people in the form. For Person 1, there is a small but statistically significant improvement in the completeness of sex reporting, but for Persons 2-6 there is a large, significant loss of information in the experimental panel.

Table 3. Percentage of Cases Missing Sex Information, by Experimental Treatment

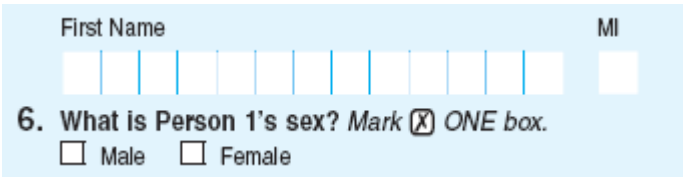
	Control	"Space Saving" Experimental (10)	t_{E-C}
Person 1	**0.68%** (.0714)	**0.47%** (.0580)	-2.26*
Persons 2-6	**1.29%** (.0925)	**3.29%** (.1504)	11.60**

*p<.05 **p<.01

These results, particularly for Person 1, may be affected by other changes in the experimental questionnaire. The small improvement for Person 1 is likely due to changes in the preceding name item. As can be seen below, adding a line for "Middle Initial" has the effect of making the sex item more visible in the experimental form:



By comparison, the write-in spaces for first name and MI form a visual barrier prior to question 6 in the control form:



Thus, requesting middle initial for Person 1 using the experimental format is likely to improve sex reporting for Person 1.

On the other hand, results demonstrate that the experimental changes led to loss of data for Persons 2-6. This may have occurred because shrinking the item to just one line led some respondents to overlook the entire item, or it may be that placing categories to the right, out of the vertical navigational path, led them to be skipped by respondents.

It might be thought that dropping the "Mark ONE…" instruction would lead more respondents to mark both response categories, but this was not the case. About 0.05 percent checked both "male" and "female" in each form.

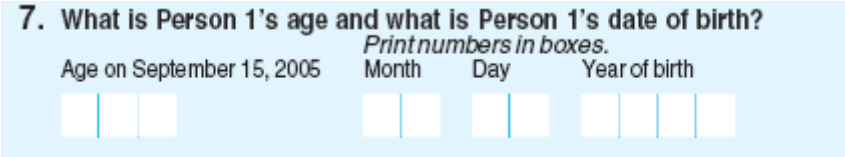
iii. Recommendation

Retain the control version of the question and instruction.

c. AGE—FORMAT

i. Experimental Treatment

The age question combines two questions into one, making for an awkward and wordy compound question in the control form:



In the experimental form, the question was shortened by turning it into an instruction. Consistent with other formatting changes made to the form, the write-in spaces were placed adjacent to rather than beneath the labels, and one of the labels ("Month") was expanded to "Month of birth." Age and date of birth are placed vertically rather than horizontally. In the experimental form, the instruction to "*Print numbers…*" is placed on the same line as the question, which seems less awkward than beneath and halfway across the line, as it appears in the control.

ii. Results

Table 4 shows results separately for day of birth, month of birth, year of birth, and age. Invalid responses are out of range (over 125 for age, values other than 1-12 for month of birth, a year of birth prior to 1879 or after 2005, and values other than 1-31 for day of birth). Responses are missing if the item is left blank.

Spencer and Perkins (1998) note that a very common error is respondents providing the last 2 digits of their birth year rather than four digits (e.g., "47" rather than "1947"). They note that this mistake can be corrected with straightforward assignment of century of birth. Although correcting two digit birth years is not quite as straightforward in the 2010 census as it was in prior censuses, it appears that most such reports can be readily corrected.[2] These errors are shown separately from other invalid year of birth reports

The formatting changes had mixed effects, leading to significant improvements for the date of birth items, but deterioration in reported age.

[2] In the 2010 census, two-digit years between 00 and 10 are ambiguous—e.g., "09" may refer to 1909, or to 2009. Age is needed to disambiguate such responses. For 94% of people for whom two-digit birth years are given, age is reported, so it is valid to assume that two digit years can be corrected in almost all cases.

Table 4. Completeness and Validity of Responses to Individual Age and DOB Items, by Format (Persons 1-6)

Item	Control Panel	"Space saving" Panel 6	t_{E-C}
Day of Birth—total	100.00%	100.00%	
Valid	**97.82** (.1228)	**98.33** (.1055)	3.22**
Invalid	0.02 (.0072)	0.02 (.0076)	.34
Missing	**2.16** (.1225)	**1.65** (.1055)	-3.23**
Month of Birth—total	100.00%	100.00%	
Valid	**97.97** (.1188)	**98.40** (.1017)	2.77**
Invalid	**0.02** (.0066)	**0.09** (0.0175)	3.76**
Missing	**2.01** (.1187)	**1.51** (.0996)	-3.26**
Year of Birth—total	100.00%	100.00%	
Valid	97.05 (.1402)	97.05 (.1316)	-.03
Two-digit birth year	**1.05** (.0800)	**1.44** (.0991)	3.07**
Invalid	0.05 (.0119)	0.03 (.0107)	-.78
Missing	**1.85** (.1121)	**1.48** (.0929)	-2.63**
Reported Age—total	100.00%	100.00%	
Valid	**97.94%** (.1119)	**96.99%** (.1267)	-5.74**
Invalid	**<0.01** (.0018)	**.03** (.0099)	2.57**
Missing	**2.06** (.1119)	**2.98** (.1263)	5.58**

** p<.01

Valid reports of month and day of birth increased by about 0.5 percentage points in the experimental form, due mostly to lower item nonresponse rates for both items.
The format changes led to less missing data for year of birth, but an increase in the fraction reporting 2-digit years. Assuming that 2-digit years are corrected, there is a significant increase in valid year of birth reporting.

These improvements are offset by a significant deterioration in the quality of reported age in the experimental form. The fraction of valid age reports declined by over 0.9 percentage points, due primarily to more missing data in the experimental form.

Table 5 shows that the net effect of these changes washes out, with no significant form difference in the overall completeness of age reporting. However, if year of birth is adjusted to treat 2-digit years as valid, then the fractions for which age can be calculated from date of birth increases to 97.64 percent for the control and 98.11 percent for the experimental panels; the difference between forms is significant ($t = 2.82$, $p<.01$). Thus the experimental changes did result in an overall improvement in age and date of birth data.

Table 5. Overall Completeness of Age Reporting, by Experimental Panel (Persons 1-6)

	Control	Experimental	t_{E-C}
Age calculated from complete DOB (Year of birth uncorrected)	96.60% (.1498)	96.67% (.1427)	.36
DOB missing/invalid, age item complete	2.41 (.1255)	2.41 (.1245)	.02
Missing or invalid responses to both age, DOB	0.99 (.0762)	0.92 (.0724)	-.75

iii. Discussion

The experimental form yields a significant improvement in calculated age, assuming that 2-digit birth years are corrected. However, the multiple formatting changes for age make it difficult to draw inferences about the cause of the positive effects on date of birth, or the negative ones on reported age, and thus difficult to support any clearcut recommendations.

The deterioration in reported age in the experimental form may reflect the combined effects of rewording the question, and format changes. Possibly, the old format made it easier for respondents to navigate and report age, since the part of the question that pertains to age appears to the left and directly above the "Age on September 15" label and write-in space, creating a vertical navigational path. It is possible that, in the control form, some respondents stop reading the question at the end of the first part of the combined question ('What is this person's age") and proceed directly to provide their age. The effect would be better age data, but worse date of birth data, as we find.

Similarly, the more complete reporting for month, day, and year of birth items in the experimental form may be due to their placement vertically below age, in the navigational path, rather than off to the right, as in the control form. The label "Month of birth" may also have helped clarify what information was to be provided.

Dillman (personal communication) advises that the performance of age in the revised format might be improved by moving the age answer space to the left, so it is placed directly above the month of birth answer space. This might be accomplished by breaking "Age on September 15, 2005" into two lines, or by abbreviating "September", e.g.

Age on Sept. 15, 2005 __ __ __

Month of birth __ __

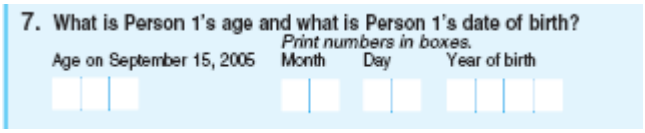
iv. Recommendations

- None

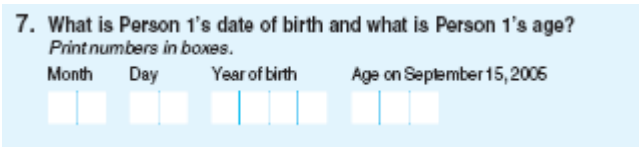
d. AGE-ORDER

i. Experimental Treatment

The traditional sequence of age and date of birth as it appears in the control form is:



In panel 4, the sequence is reversed. Panel 4 is consistent with the sequence used in the automated instruments for nonresponse followup, coverage followup, and other decennial operations.



ii. Results

Results are shown in Tables 6-9. (For this analysis, all people on the form are included, since the item reversal was carried into the continuation roster for Persons 7-12 as well.)

1. Individual Item Reporting

Table 6 shows that asking date of birth first led to improvements in all the date of birth items, with no significant effects on reported age.

Table 6. Completeness and Validity of Responses to Individual Items, by Item Order (Persons 1-12)

Item	Control Panel (Age before DOB)	Experimental Panel 4 (DOB before Age)	t_{E-C}
Day of Birth—total	100.00%	100.00%	
Valid (1-31)	**97.74%**	**98.38%**	3.80**
	(.1299)	(.1085)	
Invalid	0.02	0.01	-.58
	(.0073)	(.0061)	
Missing	**2.24**	**1.61**	-3.77**
	(.1296)	(.1085)	
Month of Birth—total	100.00%	100.00%	
Valid (1-12)	**97.90%**	**98.48%**	3.63**
	(.1258)	(.1038)	
Invalid	**0.02**	**0.05**	2.00*
	(.0072)	(.0143)	
Missing	**2.08**	**1.47**	-3.83**
	(.1257)	(.1028)	
Year of Birth—total	100.00%	100.00%	
Valid (1879-2005)	96.98%	97.27%	1.44
	(.1478)	(.1362)	
2 digit year	1.04	1.22	1.53
	(.0797)	(.0895)	
Other invalid reports	0.05	0.05	.05
	(.0118)	(.0129)	
Missing	**1.93**	**1.46**	-3.02**
	(.1194)	(.0983)	
Reported Age—total	100.00%	100.00%	
Valid (0-125)	97.86%	97.73	-.81
	(.1190)	(.1161)	
Invalid	0.01	0.02	1.15
	(.0042)	(.0072)	
Missing	2.13	2.26	.75
	(.1190)	(.1158)	

** p<.01 *p<.05

Note: Figures do not add to 100 due to rounding.

Results in Table 6 support the following conclusions:

- Day, month, and year of birth are significantly less likely to be left blank when date of birth is asked before age.
- If results for year of birth are adjusted for 2-digit birth years, then year of birth has significantly more valid reports when date of birth is asked first. (98.49 percent compared to 98.02 percent for the control; $t = 2.98$ and $p<.01$).
- There is no statistically significant effect of item order on the completeness or validity of reported age.

2. Overall Data Completeness.

Table 7 shows the completeness of age data by panel, combining information from all four items. Row 1 shows that complete date of birth was more often provided in the experimental panel. When date of birth was missing in whole or in part, information from the age item is available for slightly over 2 percent of cases in both panels. For about 1 percent of cases, both age and date of birth are missing, with no significant difference between panels.

Table 7. Overall Completeness of Age Reporting, by Item Order

	Control (Age first)	Experimental (DOB first)	t_{E-C}
Age calculated from complete DOB (year of birth uncorrected for 2-digit years)	**96.52%** (.1572)	**96.89%** (.1446)	1.81*
DOB missing/invalid, age item complete	2.43 (.1266)	2.15 (.1165)	-1.63
Missing or invalid responses to both age, DOB	1.06 (.0869)	0.95 (.0804)	-.84
Total	100.00%	100.00%	

*p<.10

Note: Figures do not add to 100 due to rounding.

When year of birth results are adjusted to correct for 2-digit year of birth reports, then age can be calculated from complete date of birth for 97.56 percent of cases in the control and 98.12 percent in the experimental ($t = 3.26$, $p<.01$).

3. Consistency of Age Reporting.

Table 8 shows the consistency between age as reported, and age calculated from date of birth, when responses to all items were present. There are no significant differences between panels in the consistency of reported age and date of birth. (These figures are not adjusted for two-digit birth years, of which there are more in the experimental form. If they were adjusted, consistency in the experimental form would improve.)

Table 8. Consistency of Age Reporting

	Control (Age first)	Experimental (DOB first)	t_{E-C}
Reported age exactly equals calculated age	89.56%	89.43%	-.43
	(.2100)	(.2235)	
Reported age is within a year of calculated age	4.26	4.50	1.33
	(.1210)	(.1271)	
Difference is greater than 1 year	1.62	1.66	.35
	(.0694)	(.0733)	
Missing	4.56	4.42	-.59
	(.1723)	(.1706)	
Total	100.00	100.00	

Note: Figures do not add to 100 due to rounding**.**

4. Age Distributions.

The data were checked to determine whether the reversed order affected age distributions. (Age was calculated from date of birth, and if date of birth was missing or incomplete, from reported age.) Table 9 shows that the same fractions report in each 10-year age group in both panels. No panel comparisons are statistically significant.

Table 9. Age Distributions, by Panel

Ten-Year Age Group	Control (Age first)	Experimental (DOB first)	t_{E-C}
0-9	10.28%	10.45%	.50
	(.2186)	(.2260)	
10-19	12.82	12.53	-.92
	(.2202)	(.2174)	
20-29	9.95	10.29	1.20
	(.2122)	(.2074)	
30-39	10.54	10.95	1.43
	(.1999)	(.2065)	
40-49	15.82	15.82	.001
	(.2304)	(.2244)	
50-59	15.68	15.16	-1.47
	(.2502)	(.2505)	
60-69	10.65	10.73	.26
	(.2139)	(.1999)	
70-79	8.22	8.15	-.24
	(.1922)	(.1988)	
80-89	4.30	4.37	.36
	(.1282)	(.1338)	
90-99	0.65	0.57	-1.25
	(.0457)	(.0440)	
100-125	0.04	0.03	-.65
	(.0104)	(.0092)	
Both date of birth and age missing	1.06	0.95	-.84
	(.0869)	(.0804)	

5. Comparison with Previous Findings.

When a reversed order of these two items was tested in the 1996 National Content Test, the results favored the age-first sequence (Spencer and Perkins 1998). Item nonresponse rates for age and for month of birth were significantly reduced when age appeared first.

Table 10 shows the fraction of people providing usable answers to the age or year of birth questions in the 1996 and the 2005 experiments, by item order. (Year of birth is the critical part of the birthdate question, according to Spencer.)

Table 10. Completeness of Reported Age and Birth Year (Adjusted) in 1996 and 2005 Experiments, by Item Order

	1996 NCT		2005 NCT	
Item Order	DOB first	Age first	DOB first	Age first
Data Completeness				
Age and birth year reported and valid	96.39%	97.17%	97.05% (.1345)	96.86% (.1448)
Age reported, birth year missing/invalid	0.26%	0.46%	**0.67%** (.0649)	**1.00%** (.0888)
Age missing/invalid, birth year reported	1.78%	1.25%	**1.43%** (.0859)	**1.16%** (.0802)
Both missing or invalid	1.58%	1.11%	0.84% (.0772)	0.97% (.0829)

Source for 1996 results: Spencer and Perkins (1998) Table 2. Results in both years are corrected for 2-digit birth years.

In the 2005 NCT, the DOB-first order was significantly more likely to be missing age with birth year present, and less likely to be missing birth year with age present. Otherwise, there are no significant panel differences in 2005 in patterns of data completeness for these two items.

Direct comparisons with the 1996 NCT are not possible, and Spencer and Perkins did not report which of the comparisons for the 1996 NCT were statistically significant.

In 1996, the age-first panel had significantly lower missing data rates for age and month of birth, which is not consistent with results shown in Table 6. The reasons for the inconsistency are not known, but may reflect questionnaire design differences. The age/DOB item appeared in a different question context in 1996 (in fact, the Spencer-Perkins analysis combines several DOB-first panels, which themselves vary in the placement and context of the item).

Finally, there is an important difference in the format of the item itself. In 1996, age and DOB were arranged vertically, with one item above the other, in both panels. In 2005, age and date of birth are presented horizontally (see facsimile on page 17). This difference may have affected item completeness, apart from any effects of the item reversal. Note that the

fraction of cases with both items missing or invalid appears higher in both panels of the 1996 NCT than in either panel of the 2005 NCT.

iii. Recommendation

- Adopt the experimental (date-of-birth-first) order.

Placing date of birth first obtained significantly more complete date of birth information, with no adverse effect on age reporting.

The experimental order improves age/DOB data, and would increase the fraction of mail households for which exact date of birth is available for matching, to identify duplicate enumerations in the census. It also would create more consistency among decennial instruments across modes.

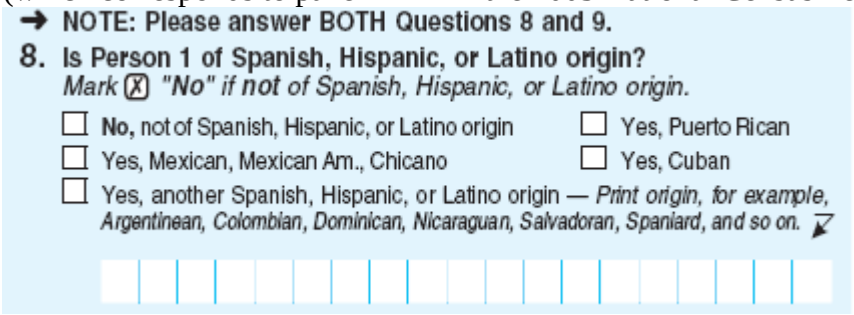
Note that another revision to this item (an instruction about how to report babies' ages) also results in improvements in age data quality and is recommended for adoption (Rothhaas et al., 2006). The reversed order recommended here has not been tested in combination with the new instruction about babies' age.

e. HISPANIC ORIGIN

i. Experimental Treatment

See Alberti (2006) for the complete analysis of the experimental test of alternative versions of race and Hispanic origin questions.

The Hispanic origin and race items are seen as redundant by many respondents. Previous research indicates that response to Hispanic origin is improved by instructing respondents to answer both questions, or by telling them to mark "no" if they are not Hispanic (Bates et al., 1995)[3]. Both instructions appear in the Control version of the Hispanic origin question (which corresponds to panel RH-4 in the 2003 National Census Test):



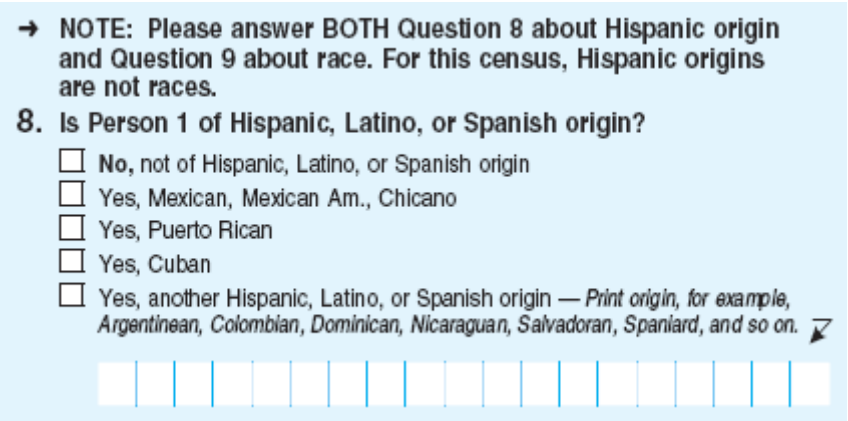
[3]I know of only one clean experimental test of the "Mark 'no' if not Hispanic" instruction; it showed a substantial reduction in item nonresponse. There is no clean evaluation of the Note instructing respondents to "answer both," although there is (confounded) experimental evidence to support its use, and it appears to function in the same way as the "Mark no" instruction. (See Bates et al., 1995, for a review of experimental evidence from Census Bureau tests.) None of those tests provide a controlled test of the effects of the Note and the "Mark no…" instruction, separately and in combination.

The 2003 National Census Test experimentally tested a stronger instruction preceding both items to further reduce item nonresponse as well as Some other race reporting by Hispanics. In the test, the instruction was expanded to say, "Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. In this survey, Hispanic origin is considered different from race. Please give different responses to Questions 8 and 9." An explanatory note was added to the race item that said, "People of Hispanic origin may be of any race." The objectives were to draw a distinction between the two items to encourage reporting to both, and to discourage reporting of Hispanic origins in the race item.

The results were mixed. Race reporting by Hispanics was more complete, and fewer reported as Some other race. On the other hand, item nonresponse increased for Hispanic origin and (among non-Hispanics) for race (Martin et al., 2004).

Results were promising enough to suggest testing a modified note in the 2005 NCT. The note was revised and cognitively tested with a small number of respondents (Hunter and de la Puente, 2005). The results indicated that some respondents still did not understand the purpose of the note. Several thought that the intent was for them to provide more specific responses than "Hispanic" by specifying a country of origin. Some thought the note could be offensive.

As part of the space-saving/simplification manipulation in panel 6, the "mark no" instruction was deleted, freeing up space for the expanded Note and for single rather than double banking the response categories, to improve their visibility. The experimental panel also reordered the descriptors "Spanish, Hispanic, or Latino" to place Hispanic first and Spanish last. This order was recommended based on cognitive testing (Gerber and Crowley, 2005).



ii. Results

Table 11 shows the frequency with which respondents marked each of the check-box categories for Hispanic origin in the experimental and control questionnaires. Responses to the write-in are not coded, and these results do not represent final Hispanic origin response distributions. These results do provide a direct measure of the effects of the instructional and design changes on check-box responses, and are suitable for the limited purpose of this analysis. (See Alberti, 2006 for the full analysis of the experimental race and Hispanic origin panels.)

Table 11. Hispanic Origin Check-Box Responses (uncoded data), by Experimental Treatment

	Control	Experimental Panel 6 ("Space Saving" changes + modified instruction)	t_{E-C}
"Not Hispanic" check-box marked	**87.62%** (.2871)	**85.27%** (.3116)	-5.86**
Hispanic origin check-box(es) marked[1]	**9.14** (.2532)	**9.83** (.2540)	2.07*
Write-in, no box marked	0.47 (.0495)	0.57 (.0619)	1.22
Missing	**2.77** (.1256)	**4.33** (.1728)	7.45**
Total	100.00	100.00	

[1]*Note:* This row includes a small number of cases in which both a Hispanic checkbox and "not Hispanic" were checked. Alberti (2006) and Census 2000 treat such answers as nonresponses.
* p<.01 **p<.01

The results show that the experimental treatment obtained more (by 1.56 percentage points) missing data. The percentage checking one of the Hispanic boxes increased by 0.69 percentage points. The percentage checking "Not Hispanic" decreased by 2.35 percentage points.

The various design changes had mixed effects; unfortunately, the confounding in this panel makes it impossible to sort them out. The decline in the fraction checking non-Hispanic, and the increase in missing data, probably are due to the elimination of the "mark 'no' if not Hispanic" instruction, based on past research. Previous experiments (Bates et al. 1995) found that deleting the instruction led more non-Hispanics to skip the question.

The statement, "For this census, Hispanic origins are not races," in the preceding note may also have contributed to item nonresponse for Hispanic origin. In recent cognitive interviews of the nonresponse instrument that I observed, this statement was meaningless to, or misinterpreted by, several respondents (all of them Black), who were stumped by it. Several respondents misheard the statement as, "Hispanics are not racists." In another cognitive test with Hispanic respondents, the statement was misinterpreted by one respondent to mean that the race question should not be answered by Hispanics. A previous version of this instruction ("People of Hispanic origin may be of any race") that was embedded in the race question was also not understand as intended, as implying a conceptual distinction between ethnicity and race (Gerber and Crowley, 2005).

The increase of 0.69 percentage points in the fraction checking a Hispanic check box may occur because the single-banked layout of the Hispanic response categories is more visible, leading more people to notice and check one of them.

If double-banking the Hispanic categories leads some to be overlooked, then we should observe differences between forms in the fraction checking the Mexican, Puerto Rican, Cuban, and other Hispanic origin check boxes. In particular, we might expect more people to mark Puerto Rican or Cuban in the experimental form, since they are placed directly below the other categories rather than off to the right. Table 12 shows that a larger fraction of people checked the Mexican box in the experimental form. There are no significant form differences for the other Hispanic origin groups (although the difference for Cubans approaches significance).

Table 12. Check-Box Responses to Hispanic Origin (uncoded data)

	Control	Experimental	$t_{E\text{-}C}$
Mexican	**5.27%**	**5.90%**	2.25*
	(.1982)	(.2171)	
Puerto Rican	1.04	1.05	.03
	(.0864)	(.0776)	
Cuban	0.36	0.47	1.60
	(.0441)	(.0539)	
Other Hispanic origin	2.28	2.16	-.66
	(.1312)	(.1182)	
Multiple boxes	0.19	0.26	1.45
checked	(.0304)	(.0375)	
Not Hispanic	**87.62**	**85.27**	-5.86**
	(.2871)	(.3116)	
No box checked	**3.24**	**4.89**	7.24**
	(.1353)	(.1880)	
Total	100.00	100.00	

*p<.05 **p<.01

These results suggest that double-banking the categories in this question may induce respondents to read them from left to right, while in the experimental form they can only be read from top to bottom. If a respondent reads from left to right, then the arrangement of Not Hispanic and Puerto Rican in the first row is easy to find, but Mexican and Cuban in the second row may be overlooked. This is somewhat speculative, but would explain why Mexican (and perhaps Cuban) is less frequently reported in the control form. Further research is needed to understand the effects of banking on respondents' reading and response behavior.

Another possibility is that reordering "Spanish, Hispanic, or Latino origin" to "Hispanic, Latino, or Spanish origin" in the question stem altered the scope of the question and the fraction identifying as Hispanic. Cognitive testing by Gerber and Crowley (2005) found that the three terms are not equivalent, either for Hispanic or non-Hispanic respondents. According to their research, "Hispanic" provides the most meaningful cue. "Latino" was sometimes interpreted as encompassing all countries with "Latin" language roots, including Italians, French, and Romanians. Most respondents identified "Spanish" with people actually born in Spain, but "Spanish origin" can imply a broader meaning, as referring to anyone with colonial ancestors in Spain. This interpretation is close to the intended meaning of the question.

In order to examine whether the experimental effects were concentrated in particular race groups, Table 12 was retabulated for each major race group (as determined by check box responses to the race question), except NHOPI, which had too few cases. The results (not shown) suggest the following conclusions:

- An increase in nonresponse in the experimental form was significant for Whites, Asians, and (especially) for Blacks. In the latter group, item nonresponse nearly doubled, from about 7 to 13 percent. This is consistent with previous research showing that the effects of the "Mark no" instruction varied by race (Bates et al., 1995).
- A decline in the fraction checking "non-Hispanic" combined with an increase in the fraction checking a Hispanic checkbox was statistically significant for Whites, for American Indians, and for Asians. For these race groups, it appears that the experimental form led more people to mark a Hispanic checkbox.
- Some other race reporters (over 70 percent of whom also checked a Hispanic box) were unaffected by the questionnaire changes.

Thus, it appears that Hispanics are largely unaffected by the questionnaire changes, but that non-Hispanics are more likely to skip the item.

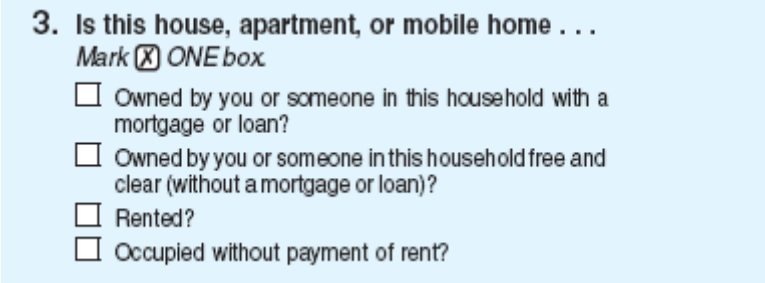
iii. Conclusions and Recommendations

- The expanded version of the Note preceding both items did not improve response. Expanding the instruction hurt response to the Hispanic origin item in the 2003 test and perhaps this one as well. Whether this reflects the content of the instruction, or its increased length, is not known. The cumulative results of these two tests suggest that efforts to instruct respondents about the distinction between race and Hispanic origin may not be worthwhile.
- Research on the effect of single and double-banking response categories should be conducted. Single banking Hispanic origin categories may improve reporting of Hispanic origin. Although the cause of the increased numbers of people checking "Mexican" in the experimental form is unknown, it is plausible that the greater visibility of single-banked categories contributed to their increased use by respondents. While the results do not prove that single-banking led to a change in reporting, there is no reason or evidence to suggest any possible negative consequences of single-banking if it is possible. Single-banked categories are generally accepted as providing a better visual display. More research on this topic is needed.
- It is not possible to conclude anything about the effects of reordering the descriptors "Hispanic, Latino, or Spanish" from this test due to confounding. Since cognitive testing indicated that this order communicated the concept effectively, the experimental order appears preferable. Note, however, that the reordering may expand respondents' understanding of the scope of "Hispanic," or contribute in some other unknown way to more people checking one or more Hispanic boxes in the experimental form.
- An experiment should be conducted in the 2010 experimental program to isolate the effects of instructional, formatting, and wording variations on Hispanic reporting.

f. TENURE

i. Experimental Treatment

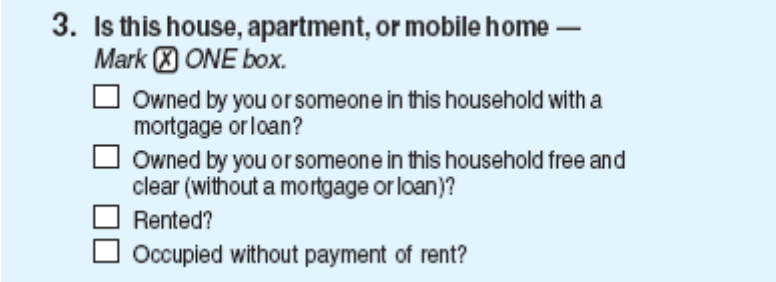
Experimental panel 6 made a very small change in tenure: it replaced the "em" dash (—) with ellipses following the question stem. The punctuation change was recommended by Dillman, Mahon-Taft, and Parsons (2004) on the basis of cognitive testing that showed respondents tended to misread the question. Respondents' confusion was resolved by rereading the question, or its categories. After the dash was replaced with ellipses, this problem did not occur in cognitive tests (Hunter and de la Puente, 2005). The experimental panel also eliminated "cash" from "rented for cash rent."



3. **Is this house, apartment, or mobile home . . .**
Mark (X) ONE box.

- ☐ Owned by you or someone in this household with a mortgage or loan?
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

The control is panel 2 (for DC-1B), which also eliminated "cash":



3. **Is this house, apartment, or mobile home —**
Mark (X) ONE box.

- ☐ Owned by you or someone in this household with a mortgage or loan?
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

ii. Results

Table 13 shows that the change did not affect the distribution of responses to the item. Item nonresponse rates are the same.

Table 13. Tenure responses, by Experimental Treatment

	Control (Panel 2)	Experimental (Panel 6)	t_{E-C}
Owned with a mortgage	49.25%	49.00%	-.42
	(.4419)	(.4250)	
Owned free and clear	24.71	25.26	.98
	(.4275)	(.3819)	
Rented	22.77	22.52	-.50
	(.3766)	(.3733)	
Occupied without paying rent	1.54	1.61	.51
	(.1102)	(.1097)	
Multiple boxes checked	0.35	0.28	-.92
	(.0582)	(.0461)	
Missing	1.38	1.32	-.42
	(.1014)	(.0997)	
Total	100.00	100.00	

iii. Recommendations

- None.

5. REFERENCES

Alberti, Nick. (2006) *2005 National Census Test Report: Results of the Race and Ethnicity Questions Analysis*.

Bates, Nancy, Martin, Elizabeth A., DeMaio, Theresa J., and de la Puente, Manuel. (1995) "Questionnaire Effects on Measurements of Race and Spanish Origin," *Journal of Official Statistics* 11(4):433-459.

Bentley, M. (2005) "Sample Specifications for the 2005 National Census Test." PRED TXE/2010 Memorandum Series: DCO-NCT2005-S-01. March 9, 2005.

Dillman, Don A., Mahon-Taft, Taj, Parsons, Nicholas. (2004) *Will a Change in Appearance Influence Responses to a Decennial Census Replacement Form Mailing Package?* Technical Report 04-040, prepared for the U.S Census Bureau.

Gerber, Eleanor, and Crowley, Melinda. (2005) *Report on Cognitive Testing of a Shortened Sequence of Hispanic Origin, Race, and Modified Ancestry Questions: Content Development for the 2005 National Content Test.* Statistical Research Division. October 18, 2005.

Hunter, Jennifer, and de la Puente, Manuel. (2005) *Report on the Cognitive Testing of the Space Saving Features, Roster Questions, and Revised Instructions for the 2005 National Census Test.* Survey Methodology Study Series #2005-03. Statistical Research Division, U. S. Census Bureau.

Lynch, Maureen P. (2005) *Review of Matches from the 2000 Acuracy and Coverage Evaluation by Matching Variables.* DSSD 2005 Memorandum Series #B-02, U.S. Census Bureau, March 7, 2005.

Martin, Elizabeth, Sheppard, David, Bentley, Michael, and Bennett, Claudette. (2004) *Results of 2003 National Census Test of Race and Hispanic Questions*, U. S. Census Bureau, October 1, 2003.

Rothhaas, Cynthia, Machowski, Jason, Tancreto, Jennifer Guarino, and Hill, Joan M. (2006) *2005 National Census Test: Tenure, Relationship, and Age Report.* U. S. Census Bureau, 2005 National Census Test Report. August 30, 2006.

Spencer, Gregory, and Perkins, R. Colby. (1998) *The Effect of Different Question Formats on Age and Birthdate Statistics from the 1996 National Content Survey*. Unpublished report, U. S. Census Bureau, January 23, 1998.

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

ACKNOWLEDGMENTS

Thanks to Florence Abramson, Theresa DeMaio, Louisa Miller, Sherri Norris, and Cynthia Rothhaas for helpful comments and corrections, and to Hector Merced for assistance with formatting.