RESEARCH REPORT SERIES
(*Survey Methodology #2005-04)*

Results of Recent Methodological Research on the Hispanic Origin and Race Questions

Elizabeth Martin[1] and Eleanor Gerber

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: August 3, 2005

---

[1]Director's Office

Results of Recent Methodological Research on the Hispanic Origin and Race Questions

Elizabeth Martin and Eleanor Gerber
U. S. Census Bureau

Paper prepared for presentation at the annual meeting
of the Population Association of America, Boston, MA[1]
April 2, 2004

Several recent developments raise concerns about the quality and consistency of racial statistics produced by government surveys.

- In 1997, the Office of Management and Budget (OMB) introduced significant changes in official racial classifications, the most important being to allow respondents to report one or more races.
- Immigration (particularly of people from Central and South America) over the past three decades has greatly increased the fraction of population which does not report in any of the major OMB race categories.
- Many surveys are collected by multiple modes, and it has proven difficult to adapt the questions on race and origin to produce comparable data in different modes.
- Interviewer administration of the same or similar questions may vary in ways that substantially influence classifications.
- Question wording, order, and other questionnaire design features influence race and Hispanic reporting.

We discuss several methodological problems that affect measurement of race and origin, and report on results of methodological research conducted to understand the impact of the questionnaire on measurements and to design better questions and procedures that will yield high quality and comparable measurements. First, we provide background for our methodological discussion by discussing the social constructs of race and Hispanic origin. We briefly describe the sources of data that we draw on, and apply them to discuss several methodological difficulties and the design solutions that have been developed to try to solve them.

RACE AND ORIGIN AS SOCIAL CONSTRUCTS

OMB has established the general principle that "self-identification is the preferred means of obtaining information about an individual's race and ethnicity, except in instances where observer identification is more practical (e.g., completing a death certificate)" (OMB, 1997:58785). The OMB standards "do *not* establish criteria or qualifications...to be used in determining a particular individual's racial or ethnic classification" and "do *not* tell an individual who he or she is, or specify how an individual should classify himself or herself" (OMB, 1997:58785).

Since 1977, OMB has held that race should be considered a social construct: "The racial and ethnic

---

[1]This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau. Thanks to Henry Woltman for advice on weighting issues in the Census Quality Survey.

categories set forth in the standards should not be interpreted as being primarily biological or genetic in reference. Race and ethnicity may be thought of in terms of social and cultural characteristics as well as ancestry" (OMB, 1997:58782). The standard defines the race categories as follows:

> *"American Indian or Alaska Native.* A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.
>
> *Asian*. A person having origins in any of the original peoples of the Far East, Southeast Asian, or the Indian subcontinent...
>
> *Black or African American.* A person having origins in any of the black racial groups of Africa...
>
> *Native Hawaiian or Other Pacific Islander*. A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
>
> *White.* A person having origins in any of the original peoples of Europe, the Middle East, or North Africa."

Hispanic or Latino ethnicity is defined as "a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race." (OMB, 1997:58789) Race and ethnicity are regarded as separate constructs (thus, Hispanic people may be of any race).

The standard also specifies certain aspects of how race and ethnic data are to be collected and reported. In 1997, the OMB required that "a method for reporting more than one race should be adopted" and "should take the form of multiple race responses to a single question and not a 'multiracial' category" (1997:58786).

In conformance with the revised standard, the race question and categories in Census 2000 mail questionnaires were:

What is this person's race? Mark (X) one or more races to indicate what this person considers himself/herself to be.

- White
- Black, African Am., or Negro
- American Indian or Alaska Native— *Print name of enrolled or principal tribe*
- Asian Indian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Other Asian— *Print race*
- Native Hawaiian
- Guamanian or Chamorro
- Samoan
- Other Pacific Islander— *Print race*
- Some other race— *Print race*

The American concept of race and the OMB standard. The OMB's adoption of a social model of race favors reliance on self-identification. However, the definition of specific, geographically based categories as standards for race reporting implies limitations on how free respondents are expected to be

in providing answers. Responses regarding social facts are easily considered "wrong" if there is a clear cultural standard defining them. Racial self identifications are problematic, because agreement about the definition of the phenomenon and the response categories is lacking. Lack of agreement arises from the inclusion within our society of persons with other cultural definitions of race, and (arguably) from ongoing change in the American concept of race itself.

The OMB categories essentially define race in terms of large geographical entities or continents. This system, while now in scientific disrepute, has had important cultural relevance in America since at least the 19th century (Gould, 1996). A feature of the American race system was codified as the old "one drop rule," which maintained that the non-White element of any person's background took precedence in defining the person's race (Tizard and Phoenix, 1995). This made it logically possible to assign only one race to a person, regardless of his or her specific ancestry. This element of the American race system is not common in other cultures, and may be changing in the current context. OMB's decision to allow reporting of more than one race reflects this changing climate.

Racial identity for Hispanic immigrants. Respondents who come to the United States from other cultures often bring with them understandings of race which do not fit with the system defined by the OMB categories. Systems from countries in Central and South America and the Caribbean, while not identical, tend to include terms that describe people who are neither Black, White nor Indian (in American terms). Hispanic respondents often search for a middle color term between "Black" and "White" and are confused not to encounter a term like "mestizo" (for Mexicans) or "trigueno" (for Puerto Ricans). Not finding middle categories is confusing for respondents who do not identify with any of the OMB categories (Davis et al., 2001). In addition, race may not be clearly demarcated from ethnicity in the systems with which some Hispanic respondents are familiar. Mexicans in the United States, for example, have referred to themselves as "la raza," meaning "the race," although the OMB treats "Mexican" as an ethnicity. Another difference is that the term "Hispanic" is frequently unfamiliar (since it is not Spanish and does not exist in the systems they learned in their countries of birth). When newly arrived persons adopt American categories, often the response cannot be expected to represent a stable, long-standing identity. It is therefore not surprising that Hispanic race reporting is relatively unreliable.

Reporting more than one race. Cognitive interviews with persons who have a multiracial heritage demonstrate the relevance of the principle that race is a social construct, and illustrate the complex factors that influence racial identifications. As Gerber, de la Puente, and Levin (1998) describe, some individuals who identify as multiracial are explicit in differentiating between what others may think they are based on observation, and how they think of themselves. They tend to reject the idea that a person's race is evident to observers, and explicitly deny that a person's sense of race can be deduced from appearance. There are also many respondents who are aware of having ancestors of more than one race, but who prefer to report in only one category. In cognitive interviews, such individuals often marked a single category while simultaneously telling researchers about relatives – sometimes parents or grandparents – who were of a different race than the one they were reporting. For them, their reported race was explicitly a social or ethnic identification. Considering that most African Americans, Hispanics and American Indians have either personal or historical knowledge of biological heritage from many continents, the number of people who make a conceptual distinction between biological heritage or appearance and social identification is potentially very large.

The change in the OMB standard to permit responses of one or more races represents a fundamental change in the way race is measured. It is not yet clear why some respondents choose to adopt this option, while others do not, or if respondents adopt both models at different times. Qualitative research suggests two patterns of reporting. Some individuals are aware that their ancestors (even parents or grandparents) derive from more than one continental "race," but this awareness has little effect on their reporting. They appear to think of race as a social identity, and frequently explain their response by referring to "the way I was raised" or other expressions of membership in a community.

A second response pattern is also observed in qualitative research. For some respondents, it is very important to report all of the elements in their backgrounds. This appears to arise from a psychological rather than a social approach to racial identification. The motivation seems to be to express "who they are" rather than to indicate a community membership. In some cases this reflects closeness to particular family members. Mentioning only one race seems to these respondents to exclude or deny that relative or part of the family. Others report even distant ancestries as a way of accounting for aspects of perceived identity. The psychological approach may follow a belief that is sometimes elicited in discussions of ancestry: that one's personality, character and capacities may be inherited, and that accounting for different ethnicities helps explain them.

These two approaches imply that multiracial reporting has widely different salience for different respondents. It is reasonable to assume that respondents for whom it is highly salient are among those who report more than one race even when the question does not permit it, and who refuse to offer a race with which they "most identify" in follow up questions.

The phenomenon of contextualized race reporting for multiracial individuals has been observed in other research. For example, Xie and Goyette (1997) found that Asian-White children's self-identification can be considered "optional." Other research points to the flexibility and contextuality of self-labels. Self-identification for multiracial persons may vary according to the social milieu. Phinney and Alipuria (1996) found that the self-label chosen by multiracial teenagers varied by the ethnic composition of their schools, and conclude that there is "clear evidence of variability of multiethnic student's ethnic self-labels in different contexts." They also found a relationship between the form of the question and the response: write-in mode and multiple option questions elicited different patterns of response. It has also been suggested that multiracial teenagers may offer different responses depending on the presence or absence of their parents during the interview. Harris (2002) found that adolescents were twice as likely to report multiple races in an interview conducted at school than an interview conducted at home. Both interviews asked for self-reported race, but the school interview was self-administered while the home interview was interviewer-administered and often observed by family members.

Thus, multiracial individuals may offer different accounts of their identification at different times, for complex social and psychological reasons. Questions that elicit one, stable identifier may be difficult or impossible to construct for persons with socially fluid identities. Proxy reporting might introduce additional variability, although its role has not yet been investigated. Factors associated with the design of the question (discussed below) also may contribute to poor reliability in multiracial reporting in surveys. Bennett (2003) found that the consistency of reporting varied for particular racial combinations. The extent to which the inconsistencies reflect real variation or change in racial identifications, or the influence of methodological factors is unknown. Coding and editing conventions

may also introduce variability.[2]

SOURCES OF DATA

This paper draws upon two experiments that focused on the effects of the design of the Hispanic origin and race questions on reporting, and one reinterview study conducted a year after the census.

The Census 2000 Alternative Questionnaire Experiment (AQE) was conducted during Census 2000, and included two experimental treatments. In one, 1990-style short forms were mailed to a random sample of 10,500 households. The 1990-style form preserved 1990 census question wording, categories, order, and format, but incorporated some recognizable elements of the 2000 design. A second, control panel of about 25,000 households received Census 2000 mail short form questionnaires. The purpose was to evaluate the effects of changes in the design of the questionnaire on reporting of race and Hispanic origin, and to assess the implications for the comparability of race and origin data collected in the 1990 and 2000 censuses. Mail return rates were about 72 percent for both panels. For respondents in the AQE, the responses provided on the mail forms were their census data. Households which did not return a mail questionnaire were followed up as part of the normal Census 2000 nonresponse operation. They are not included in this analysis. All experimental data were keyed and processed separately from the production census. Data for both forms were edited by applying a simplified version of the pre-edits used in Census 2000 production. Missing data were not imputed or allocated, as they would be in fully edited census data. See Martin (2002) for details.

The 2003 National Census Test (NCT) sampled about 160,000 households from a universe of Census 2000 mailout-mailback areas, representing about 97 percent of the country. Seven experimental panels varied four different features of the questionnaire design: (1) wording changes to the Hispanic origin item, (2) dropping the Some other race category from the race item, (3) modified instructions, and (4) examples in the race and Hispanic origin items. The design of the experiment is presented in an Appendix. Each panel was allocated a sample of 20,000 housing units. The response rate was about 67 percent for the test. The estimates from the sample have been weighted and represent about 100 million households. See Sheppard, Bentley, and Woltman (2003) for details.

Both NCT and AQE samples excluded households outside of Census 2000 mailout/mailback areas, and therefore, results can be generalized only to these areas. This excludes several important groups that are not enumerated by mailed questionnaires, such as American Indians on reservations and Alaska Natives. In addition, both studies include only households that responded by mail. Characteristics of responding households, including Hispanic origin and race, are not expected to be identical to Census 2000.

The Census Quality Survey (CQS) replicated Census 2000 data collection methods in a reinterview conducted about a year after the census. In June 2001, short form questionnaires that asked the Census

[2]Some inconsistencies may be more apparent than real because of the way responses are coded. For example, a person who wrote in "biracial" would have been coded as Some other race, while the same person who wrote in or checked "Black and White" would have been coded as "Black" and "White," resulting in inconsistencies in the number of races reported as well as the race categories. In this case, different labels with similar meaning used in the two interviews caused the inconsistency, which can be addressed by refinements to the coding scheme.

2000 race and origin questions were mailed out to sample households. Housing units containing people who reported two or more races in the census were oversampled. (Results reported here are for Panel A only, with N=27,500; see Bentley et al. 2003.) About 54 percent of households responded by mail. Enumerators followed up nonresponding households and visited addresses in nonmailout/mailback areas. Unlike the census, the CQS did not permit non-household proxy respondents. Individuals' answers were matched to their census responses to examine the consistency of race reporting. Based on the CQS respondent's report, it was determined that about 59 percent of mail questionnaires were filled out by the same person in the CQS and the census. The CQS used the same coding and editing procedures as the census. Despite the comparability of the methods used in the two surveys, the consistency of reporting multiple races was very low. Only 40% of people who reported two or more races in the census also reported two or more races in the CQS interview. See Bentley et al. (2003) for details.

All data presented here are weighted. Standard errors and significance tests are calculated using jackknifed replication methods in VPLX (Fay, 1998).

QUESTIONNAIRE ISSUES

Several challenges arise in developing survey questions on race and origin that provide valid measurements of race and origin. The following have proved problematic: distinguishing race and ethnicity questions, communicating an appropriate concept of race, communicating the option of choosing more than one race category, question sensitivity, communicating the intent of the Hispanic origin question, category problems, and mode and interviewer effects.

1. Distinguishing questions about race and ethnicity.

Although the OMB defines Hispanic origin as distinct from race, many respondents do not see a difference (Gerber, de la Puente, and Levin, 1998). The perceived redundancy is a major source of reporting problems. When race is asked first in a self-administered questionnaire, many Hispanic respondents look for but do not find a category to report themselves, and so either leave the question blank or mark "Some other race" and write in a Hispanic group, such as "Mexican" or "Salvadoran" (see, e.g. Kissam, Herrera, and Nakamoto, 1993). Many non-Hispanic respondents skip Hispanic origin, apparently thinking it is redundant or does not apply to them. It is difficult to communicate a conceptual distinction that many respondents do not find valid or natural, but some questionnaire design features help reduce reporting problems.

The perceived redundancy in the 1990 census form is exacerbated because the items are sequenced with the more general item (race) preceding the more specific one (Hispanic origin). Methodological experiments show that reversing the order of the two items to ask Hispanic origin first and adding an instruction to answer both reduces the apparent redundancy and allows Hispanic respondents to report their Hispanic origin before responding to the question on race (Bates et al., 1995; U. S. Census Bureau, 1996a). On the basis of the research, these changes were implemented in the Census 2000 mail questionnaire. (The placement of the question on Hispanic origin before the question on race was also adopted as an OMB standard in 1997.)

Because of the changes, fewer respondents skipped the Hispanic origin item in Census 2000 forms. According to the AQE, Hispanic origin item nonresponse was 3.3% in 2000-style forms, compared to 14.5% in 1990-style forms (Martin, 2002). In addition, fewer Hispanics left the race item blank or reported in Some other race (SOR), as shown in Table 1. (The difference in the percentage reporting SOR is only marginally significant.) The net effect was a significantly larger number of Hispanics reporting in the OMB race categories–45% in the 2000-style forms, compared to about 32% in the 1990-style forms. Although this represented a major improvement, nonresponse was still very high, with fewer than half of Hispanics reporting in the official race categories.

Table 1. Race Reporting by Hispanics in Census 2000 Alternative Questionnaire Experiment, by Form Type

	1990-style form	2000-style form	$t_{1990\text{-}2000}$
Reported in an OMB race category	31.5% (1.90)	44.9% (1.43)	5.61*
Reported Some other race (SOR)	36.9% (2.00)	33.0% (1.40)	-1.63†
No answer	31.1% (1.93)	22.2% (1.20)	-4.12*
Total (single race reports only)	100.0%	100.0%	

* $p<.001$ †$p<.10$ (one-tailed)

The Census Bureau has continued to experiment with methods to try to clarify for respondents the distinction between race and Hispanic origin and increase reporting in the OMB categories. In the 2003 NCT, two new mail questionnaire design features were tested: strengthening the instruction, and dropping the Some other race category.

The instruction that preceded the race and origin questions in the Census 2000 mail questionnaire said, "NOTE: Please answer both Questions 5 and 6." It was revised as follows to try to better communicate the distinction between the two items: "Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. In this survey, Hispanic origin is considered different from race. Please give different responses to Questions 5 and 6." An explanatory note added to the race item said, "People of Hispanic origin may be of any race." Although the revised instruction improved Hispanic race reporting, it led to more nonresponse for the Hispanic origin item (Martin et al. 2003b). More developmental work and testing are needed on the instruction.

The purpose of dropping the category is obvious: reduce the number of people who report as Some other race and whose race responses must be imputed for many uses of census data. While this may seem an inevitable effect of dropping the category, in fact we could not anticipate how many respondents would still write in a response (e.g., "Hispanic") that would be coded as Some other race. We were also concerned about an increase in the number of Hispanics who would skip the question on race because they could not find a category for themselves.

Both anticipated effects were confirmed by the 2003 NCT. As shown in Table 2, dropping the SOR

category reduced (by 18.2 percentage points) but did not eliminate Some other race reporting by Hispanics, and increased race non-response for Hispanics (by 6.5 percentage points).

Table 2. Average Effects on Hispanic Race Reporting of Dropping the Some Other Race Category (2003 National Census Test)

	Questionnaire Design Treatment			
Hispanic race responses	With SOR category	Without SOR category	Effect of dropping SOR	t
1. Reported in an OMB race category	49.8% (1.49)	61.4% (1.44)	11.6	5.8*
2. Reported Some other race	30.9 (1.25)	12.7 (.91)	-18.2	-12.6*
3. No answer	19.3 (1.11)	25.8 (1.23)	6.5	4.2*
Total	100.0%	100.0%		

Note: Comparisons are between combined panels 3 and 4 (no SOR category) versus 1 and 2 (with SOR category); see Appendix.
*p<.001

The loss in race data for Hispanics (due to non-response) was more than offset by the gains from reduced reporting of SOR. There was a net gain of 11.6 percentage points in the fraction of Hispanics reporting in the OMB race categories.

Dropping the Some other race category also led to more complete race data for non-Hispanics (item non-response declined from 1.9 to 1.5 percent), and more complete Hispanic origin data (item non-response declined from 3.7 to 3.3 percent). These small, but statistically significant, improvements were unanticipated. Possibly, the items are less confusing when the SOR category is removed, but more research is needed to understand the basis for the effect.

Based on these and other results from the 2003 NCT (see Martin et al., 2003b; Sheppard, Bentley, and Woltman 2003), the Census Bureau's current plans are to drop the Some other race category from the question on race in the 2010 census.

2. Communicating an appropriate concept of race.

The wording of the question must be sensitive to the intent to measure social identification, as distinct from biological heritage and appearance. We find the wording of the following race item problematic:

> "I'm going to read a list of race categories. Please choose one or more categories that best describe [NAME'S] race."

"Describe" carries strong visual connotations, and is likely to strike respondents as a request for how others literally see them, placing undue emphasis on external features like hair and skin color. In our view, this negates the principle of self-identification by making the concept physically, rather than socially, based. We prefer phrases such as "consider him/herself to be" or "identify" because they clearly mark the question as referring to a subjective, rather than physical, set of facts.

Another wording problem in the question above is the term "best," which may be interpreted as asking for the race that is thought to be superior to others. This problem surfaced in early cognitive testing of a similar question, which some respondents interpreted as asking them to choose which race they thought was the best race (U. S. Census Bureau, 1996b). A follow-up question asked in the National Health Interview Survey when respondents reported more than one race ("Which one of these groups...would you say BEST represents your race?") also invites this problematic interpretation. This may have contributed to high item nonresponse rates of 23% for White/Black and 13% of White/Asian or Pacific Islander respondents (Lucas et al. 1999). In order to avoid implying invidious comparisons among race groups, we advise against using the words "best" or "better" in the question. However, we are not aware of experimental evidence on the effects of such wording differences on reporting.

3. Communicating the "one or more" option in the race question.

For most people, being asked about and reporting their race is very familiar, and this made it difficult to get them to notice the novel "one or more" option that was introduced in the 2000 census. The way respondents read questions accounts for some of the difficulty. Cognitive interviews suggest that respondents are less likely to read instructions than the question stem. Even when they read an instruction, they may not absorb its meaning. In initial cognitive testing of "one or more" questions for Census 2000, respondents often did not realize they could mark more than one category, even when they had just read the pertinent instruction aloud (Gerber, de la Puente, and Levin, 1998). The "one or more" option is often indicated with plural grammatical forms, such as "race or races" or plural verb forms which may be overlooked in reading (Baucom, 1970). They are somewhat more effective in spoken mode, and have been used in interviewer-administered surveys.

In addition, questions that ask respondents to "mark all that apply," similar to the census race question, typically do not obtain complete reports (Rasinski, Mingay, and Bradburn, 1994). Respondents may not mark all the race categories that reflect their racial heritage because the format of the question does not press them to, and variability over time in whether they in fact "mark all that apply" may contribute to unreliability.

Confusion also may be caused by contradictory syntactic elements of the question. In the following item, "one or more" is contradicted by the singular reference to "race."

> "...Please choose one or more categories that best indicate your race."

Even though "best indicate" is correctly in the plural, the question is interpretable as a request for a single race. Choosing something that is "best" implies winnowing alternatives and retaining one.

In the question used in the Census 2000 mail questionnaire, the question stem ("What is this person's race?") is singular, and the possibility of reporting one or more races is introduced only in the instruction

that follows ("Mark one or more races to indicate what this person considers himself/herself to be.") When one part of the question on race uses the singular concept and another uses a multiple concept, respondents may be unsure whether they may report more than one race. Such wording problems may contribute to the unreliable reporting of two or more races in repeated measurements of the same people, as discussed below.

Table 3 shows the consistency of race reporting by non-Hispanics reinterviewed after the 2000 census in the Census Quality Survey. The rows represent the major census race category or categories in which a person was reported in the census. The first column shows the fraction of each race group reporting exactly the same major race (or combination of major races) in the CQS. The second column shows the fraction reporting in CQS one or more, but not all, of the multiple races reported in the census. The third column shows the fraction reporting the same race(s) reported in the census, as well as additional races. Finally, the fourth column shows the fraction reporting completely different major race(s) in the CQS.

Table 3. Consistency of Major Race Reporting in Census 2000 and the Census Quality Survey, for Single OMB Race and Selected Multiple Race Categories (Non-Hispanics only)

	Race reported in CQS				
Major race reported in the census	Same race	Subset of multiple races	Same race + others	Different race(s)	Total
White	97.9% (.37)	—	1.1 (.27)	.9 (.25)	100.0%
Black	97.2% (.85)	—	1.0 (.39)	1.8 (.74)	100.0%
American Indian or Alaska Native	53.7% (21.73)	—	33.8 (25.51)	12.6 (7.34)	100.0%
Asian	95.7% (2.25)	—	2.5 (1.98)	1.8 (1.09)	100.0%
Native Hawaiian or Other Pacific Islander	56.4% (8.84)	—	37.8 (8.58)	5.8 (2.30)	100.0%
White + Black	56.1% (1.59)	28.1 (1.40)	6.0 (.78)	9.8 (1.02)	100.0%
White + Asian	55.3% (1.59)	35.5 (1.54)	2.5 (.48)	6.6 (.71)	100.0%
White + Am. Indian or Alaska Native	40.1% (1.28)	57.8 (1.28)	.6 (.16)	1.5 (.25)	100.0%
Black + Am. Indian or Alaska Native	29.4% (2.79)	60.9 (3.01)	3.1 (.88)	6.5 (1.45)	100.0%

Note: The cross-sectional weight (z_wgt1) is applied to persons reporting one race, and the bridge weight (z_wgt2) is applied to persons reporting two or more races. (See Bentley et al. 2003.)

For some single race groups (American Indian and Alaska Native, and Native Hawaiian and Other Pacific Islander) and all of the multiple race groups shown in Table 3, reporting reliability is low. At best, slightly over half report the same race(s) in CQS they had reported in the census. However, most of the inconsistency does not arise from completely different race reports, but rather from respondents reporting some but not all of the races originally reported in the census, or reporting races in addition to those reported in the census. Reports involving American Indian and Alaska Native (AIAN) race are particularly fluid in this way. Note that a third of people reporting AIAN in the census gave additional races in the CQS. And large majorities of people who reported White + AIAN or Black + AIAN in the census reported only one of the component races in the CQS.

Although consistency is improved somewhat by restricting the sample to those reports given by the same respondent in both CQS and census, the results are similar to those shown in Table 3.

It will require a good deal more research to understand the causes of the inconsistent race reporting shown in Table 3. It is an open question whether the inconsistency can be reduced by questionnaire improvements. However, some question revisions might better communicate the one or more option.

The difficulty of absorbing the option in personal or telephone interviews may be exacerbated because respondents are engaged in competing, simultaneous mental tasks. One way to communicate the "one or more" option is to probe for "Any more?" as is often done in survey questions permitting more than one answer. The problem is that the probe may be leading, and is vulnerable to interviewer effects. A behavior coding study found that interviewers in one survey asked the probe in about half of interviews, indicating a failure of standardization (Schwartz, Fricker, and Dixon, 2001). Cognitive testing suggested that one third of respondents were unaware they could choose more than one category (Fisher, Fricker, and Schwartz, 2000), so the probe did not communicate the option in a uniform way.

A partial solution, which we favor, may be to introduce redundancy by repeating the multiple response option on a flashcard or in the question, giving the respondent more than one chance to absorb it. The following question uses two sentences, with each sentence communicating the option in a slightly different way:

> "Now choose *one or more races* for each person. Which *race or races* does each person consider himself/herself to be?"

4. Question sensitivity.

As Gerber, de la Puente, and Levin (1998) discuss, respondents frequently do not understand the purposes for collecting information about race in the census, and many object to it based on their assumptions about the effects on society of asking about race. Race and ethnicity questions are often seen as part of a political and social dialogue about race, and certain response patterns make sense only in that context. Respondents who experience the questions as socially divisive may attempt to provide an inclusive answer by, for example, writing in "American" or "human."

Thus, respondents think not only about the content of the question about race, but about the possible social and political implications of its being asked, and of their answering it in particular ways. The race question is primarily sensitive because it is asked, not because the information revealed is sensitive

information. The categories also may be a source of sensitivity. Certain terms may be perceived as inappropriate, such as "Negro" (Gerber, de la Puente, and Levin, 1998). In addition, respondents in qualitative studies are highly sensitive to aspects of the question that suggest that different groups are not treated equivalently. For example, non-Hispanic respondents sometimes comment on the "unfairness" of the Hispanic origin question, and question why one group would be given "special notice." Hispanics may also object to being "singled out" by the separate question. Black respondents question why their category includes the modifier "American" (in "African Am.") while others do not. Black or White respondents may point out that other groups have the opportunity for specific reporting (available through write-in lines or detailed checkbox categories) that their group lacks. This may be perceived as unfair, and in reaction (or protest) respondents may write their ethnicity on any available write-in line. Thus, a write-in entry of (for example) "Italian/Irish" on the Asian write-in line may represent an attempt to achieve fairness, rather than a misunderstanding of the intent of the question (Gerber, de la Puente, and Levin, 1998). Finally, as noted above, a question that asks which category "best indicates" or "best describes" their race may imply invidious comparisons and cause resentment.

5. Communicating the intent of the Hispanic origin question.

As noted above, one problem with the Hispanic origin question is that non-Hispanics may not realize they are supposed to answer it. Reversing the sequence of race and origin went a long way toward reducing high rates of item nonresponse found in the 1990 census. Other changes were made to simplify the question, make it more user-friendly and use terms familiar to the broadest number of Hispanics. The term "Latino" was added to the question, the word "origin" was dropped, and the examples were dropped.

The 1990 question and categories were:

> "Is this person of Spanish/Hispanic origin? *Fill ONE circle for each person.*
> No (not Spanish/Hispanic)
> Yes, Mexican, Mexican-Am., Chicano
> Yes, Puerto Rican
> Yes, Cuban
> Yes, other Spanish/Hispanic *(Print one group, for example: Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.)*"

In 2000, the question was:

> "Is this person Spanish/Hispanic/Latino? *Mark the **'No'** box if **not** Spanish/Hispanic/Latino.*
> **No,** not Spanish/Hispanic/Latino
> Yes, Mexican, Mexican-Am., Chicano
> Yes, Puerto Rican
> Yes, Cuban
> Yes, other Spanish/Hispanic/Latino— *Print group.*"

Although the changes may seem minor, they apparently affected the quality of the data obtained in Census 2000. There was a loss of information about detailed Hispanic origins which was attributed to

the dropping of examples (Logan, 2002; Scott, 2001, the Los Angeles Times, 2001).

Results of the AQE confirmed that a 1990-style Hispanic origin question elicited more detailed reports of Hispanic origin (such as "Colombian" or "Salvadoran") than the Census 2000 question, which obtained more generic reports, such as "Hispanic," "Latino," or "Spanish" (Martin 2002). Table 4 shows questionnaire differences in reporting in four Hispanic origin groups:
(1) groups with check boxes (Mexican, Puerto Rican, Cuban) in both 1990- and 2000-style forms;
(2) groups listed as examples in the 1990-style form but not the 2000-style form (Argentinian, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard),
(3) all other specific groups without check boxes and not listed as examples in either form, and
(4) generic origins, such as "Hispanic," "Spanish," or "Latino."

Table 4. Percentage of Hispanics Reporting in Four Origin Groups in Census 2000 Alternative Questionnaire Experiment, by Form Type

	2000-style	1990-style	$t_{2000\text{-}1990}$
1. "Check box groups": Hispanic groups with separate check boxes in both forms	70.2%	73.2%	-1.37
2. "Example groups": listed as examples in 1990-style form but not Census 2000	6.4	11.2	-3.58*
3. All other detailed Hispanic groups	4.2	8.7	-3.38*
4. Generic origins (Write-in is general descriptor "Hispanic" "Latino" "Spanish")	11.9	1.9	10.32*
No write-in (or write-in uncodable)	7.2	5.0	2.15*
All people identified as Hispanic	100.0%	100.0%	
N	5,163	3,091	

*$p < .05$

Overall, about 92% of Hispanics reported a specific group in 1990-style forms, compared to 80% of those who filled out 2000-style forms. The 1990-style form elicited more detailed reporting in all three detailed categories (groups with separate check boxes, those listed as examples, and the remaining groups). In the 2000-style forms, Hispanics tended to describe their ethnicity in general rather than specific terms (category 4). About 12% wrote in Hispanic, Latino, or Spanish as their "group," compared to about 2% in 1990-style questionnaires. However, the effect did not appear to be due entirely to the presence of examples. There was a significant difference in reporting in one of the "checkbox groups" (59% reported as "Mexican" in the 1990-style forms, compared to 54% in the 2000-style questionnaire). Since the wording of this category was identical in both forms, design differences other than the examples must account for the reporting difference. Possibly, dropping the word "origin" altered the meaning of the question.

The 2003 National Census Test experimentally evaluated alternative solutions to the problem. The question wording was revised to restore the word "origin" and the examples that had been used in 1990,

as follows:

"Is this person of Spanish, Hispanic or Latino origin?
No, not of Spanish, Hispanic or Latino origin
Yes, Mexican, Mexican Am., Chicano
Yes, Puerto Rican
Yes, Cuban
Yes, another Spanish, Hispanic or Latino origin–*Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*"

The effects of these two changes were independently tested in separate panels and the results are shown in Table 5. Column (1) shows the effects of the revised question wording when neither question included examples. Restoring the term "origin" and the other minor wording changes cut the fraction reporting generic origins in half. Results presented separately for the three checkbox categories (Mexican, Puerto Rican, and Cuban) confirm the inference drawn from the AQE that the question wording reduced the fraction reporting as Mexican in the Census 2000 Hispanic origin question. The item wording did not significantly affect the fraction reporting as Puerto Rican or Cuban.

Table 5. Percentage of Hispanics Reporting in Four Origin Groups, by Questionnaire Treatment (2003 National Census Test)

	(1) No examples and Hispanic origin question has...			(2) Revised question wording		
	Census 2000 wording	Revised wording	t	Without examples	With examples	t
Checkbox groups	67.2% (1.51)	74.3% (1.54)	3.14*	73.3% (.87)	73.8% (1.02)	.4
Mexican	51.7 (1.68)	58.3 (1.73)	2.65*	56.9 (.99)	58.3 (1.14)	.96
Puerto Rican	11.4 (.97)	11.9 (.99)	.44	11.2 (.63)	10.8 (.58)	.51
Cuban	4.1 (.64)	4.1 (.60)	-.04	5.2 (.48)	4.7 (.45)	.77
Example groups	6.7 (.74)	9.1 (1.04)	1.86†	9.1 (.55)	10.2 (.63)	1.33
All other detailed Hispanic origins	7.9 (.94)	7.2 (.84)	-.57	8.4 (.54)	9.3 (.56)	1.18
Generic origins ("Hispanic," "Spanish," "Latino" or "Other" box checked, with no writein)	18.2 (1.36)	9.5 (.95)	-5.38	9.2 (.56)	6.7 (.53)	3.18*
All people identified as Hispanic	100.0%	100.0%		100.0%	100.0%	

Note: The wording effect was tested by comparing the Control and panel 1. Example effects are evaluated by comparing combined panels 1, 3, and 5 versus 2, 4, and 6. See Appendix.
*p<.05
†p<.10

Column (2) shows the effects of examples on detailed Hispanic reporting, in the context of the revised question wording. There were no significant reporting differences between panels with examples and those without, with the exception of the decline in reporting of generic origins in the presence of examples. Examples evidently helped respondents understand that the intent of the question is to obtained detailed origin information.

6. Category problems and example effects.

Although the category system mandated by OMB works well for many respondents, there is evidence that many others have difficulty understanding and choosing among the categories. The consequence is

high rates of missing race data and unreliability for some groups, especially Hispanics.

Some groups object to the general classification in which they find themselves. For example, some members of the group categorized as "Asian Indians" point out that geographically, India is not in Asia (Gerber, de la Puente and Levin, 1998). Other groups do not perceive themselves as members of the groups into which OMB has placed them. For example, Arabs may not see themselves as belonging to the White category. Other groups also fail to find a category that expresses their own sense of race, the largest being Hispanics who wish to report Hispanic (or a Hispanic nationality) as their race.

Effects of labels and examples. Particular labels affect respondents' understanding of which, if any, category they belong in. Seemingly equivalent categories may not have the same meaning. For example, "Black" has a wider scope than "African American," which applies only to persons whose ancestors have lived in the United States for a long period of time. The term "African American" was an obstacle for Black respondents from the Caribbean or Africa, who hesitated to mark the "Black, African Am., or Negro" category because they were uncertain if it was intended to apply to them (Gerber, de la Puente, and Levin, 1998).

Racial terms appear not only in the checkbox category labels, but may also be used as examples. As noted above, examples were dropped from both race and Hispanic origin questions in Census 2000.

Examples may affect the interpretation of the question, by illustrating the intended specificity of responses. This type of effect on question comprehension apparently influenced reporting in the Hispanic origin question in Census 2000, as described in section 5.

However, examples also can distort reporting. "English" appeared first in the list of examples following the ancestry question in 1980, but was dropped in 1990. There was a corresponding drop from 1980 to 1990 of about 17 million persons reporting English ancestry. There were also large increases in the numbers reporting German, Acadian/Cajun, or French-Canadian ancestry, apparently due to the listing of these ancestries as examples in 1990 but not 1980, or their greater prominence in the 1990 list. These effects of examples, and their order, may occur because respondents write in the first ancestry listed that applies to them (Scarr 1993).

The design of the questionnaire may influence how examples are interpreted and used. In the 1990 version of the race question, the "Other API" category and write-in box had the following instruction and examples printed off to the side: "If Other Asian or Pacific Islander (API), print one group, for example: Hmong, Fijian, Laotian, Thai, Tongan, Pakistani, Cambodian, and so on." In the AQE, these seven groups were significantly *less* likely to be reported in the 1990-style form, where they appeared as examples, than in the 2000-style form, where they did not (Martin et al., 2003). This result is counterintuitive and puzzling, and contrasts with the effect of examples in the Hispanic origin question. Possibly, the placement of the examples far to the left of the question, remote from the write-in space, meant that many respondents never saw the examples, while those who saw them did not associate them with the write-in space. The examples may have distracted respondents and disrupted the response process. Perhaps some respondents who found the examples circled them, without writing in a response. In the 2003 NCT, exactly the same race examples were used in the race question adjacent to the write-in space, and had very different effects. The examples resulted in a significant decline in generic Asian or Pacific Islander race reports and an increase in reporting of specific races. In this placement, the

examples apparently helped respondents understand the intent of the questions. The inconsistent effects of the same examples in the two experiments provide a caution that a questionnaire feature (such as examples) may interact with the graphical organization of the form (such as their placement) to produce different effects on the response process and the data.

7. Designing questions that are comparable in different survey modes.

The challenges of collecting comparable race and origin data that are not vulnerable to methodological differences between surveys is greater because of the difficulties of adapting questions for administration in different survey modes, and the potential influence of interviewers on measurements. It is difficult to gauge the importance of interviewer and mode effects, because the experimental research needed to disentangle their effects from the effects of respondent self-selection has not been conducted. Thus, for example, in the census Hispanics and persons whose native language is not English tend not to respond by mail, and so end up being interviewed in nonresponse followup. Apparent mode differences in race and origin reporting can result from the direct effects of mode or from differences in the populations enumerated by each mode or from inadvertent changes in the meaning of the questions as they are adapted for different modes, or all of the above.

Because the list of race categories is long, and because the categories may not correspond to some respondents' own understandings of race, it is usually thought necessary to communicate all the categories in which respondents are expected to respond.

In a self-administered questionnaire it is easy to print a long list of categories below the question. Even in this mode, respondents may not understand that the numerous categories are all part of the same question, especially since some categories include write-in spaces that visually fragment the list (as is the case in the Census 2000 mail questionnaire). In personal interviews, communicating the categories is facilitated by use of a flashcard. Telephone interviews present a greater challenge, because the flashcard is not available to communicate the categories or aid in comprehension of the "one or more" option. The question must be reworded, with greater redundancy needed than in other modes to communicate the "one or more" option. The list of response categories must be shortened or restructured because of the difficulty of communicating a long list orally. Most telephone surveys adapt the question by setting up a branching structure.

Some rewording is necessary to administer the question in different modes, but sometimes the adaptations inadvertently alter the meaning of the question. This may lead to different questions being administered in different modes in the same survey. (1) and (2) were asked in the mail questionnaire and in nonresponse followup interviews, respectively, in the 2000 census; their wordings seem close, although not identical.

(1) "What is this person's race? Mark (X) one or more races to indicate what this person considers himself/herself to be."

(2) "Now choose one or more races for each person. Which race or races does each person consider himself/herself to be?"

(3), (4), and (5) were asked in the American Community Survey/Census 2000 Supplementary Survey in

the mail questionnaire, the computer-assisted telephone (CATI) nonresponse interview, and the computer assisted personal (CAPI) nonresponse interview, respectively:

(3) "What is this person's race? Mark (X) one or more races to indicate what this person considers himself/herself to be."
(4) "I am going to read you a list of race categories. Please choose one or more categories that best indicate (Name's/your) race."
(5) (Show flashcard) "Using this list, please choose one or more categories that best indicate (Name's/your) race."

The alterations made in (4) and (5) seem to have changed the substantive meaning of question (3). Note the introduction of the term "best", which is problematic for reasons discussed above. Item (5) relies on a flashcard, which was available in CAPI but not CATI. The CATI question (4) gives respondents only one chance to absorb the "one or more" option. These wording difference s may affect the comparability of the data collected in different modes. In addition, the ACS/C2SS data were intended to be comparable to Census 2000 data. Note that, although the mail questions (1) and (3) are identical, the questions asked in nonresponse followup are different, as can be seem by comparing the wording (2) with (4) and (5).

8. Interviewer effects.

The race question poses problems for interviewers which they solve in different ways, giving rise to interviewer effects that may substantially influence the data. Although typically instructed to read the question and categories as worded, interviewers apparently vary considerably in how (and whether) they administer the question, and the degree to which they probe responses that do not fit the printed categories. In censuses before 1980, enumerators recorded race based on their own observation (U. S. Census Bureau 1983), and this practice may still sometimes occur. In the 1980 census, Hispanics were far less likely to be reported as "other race" in personal visit reinterviews than in self-administered census questionnaires (McKenney, Fernandez, and Masamura, 1985). The explanation is that many Hispanics who reported themselves as "Some other race" on their census questionnaires were recorded as "White" in subsequent reinterviews, probably based on the interviewers' observations. More recently, a behavior coding study found that interviewers made major changes to the wording of race question in over 40 percent of the personal interviews that were behavior coded, usually by omitting categories, and skipped it in 7 percent of interviews (Smiley and Keeley, 1997). Differences in training may also influence how interviewers administer the questions, and the results they obtain.

For the most part, the mode and interviewer effects on measurements of race and origin have not been systematically evaluated, although several comparisons among and within surveys suggest their effects may be substantial. For example, comparisons of race data from Census 2000 and the Census 2000 Supplementary Survey, or C2SS, reveal sizable differences in race distributions, particularly for Hispanics.

Bennett and Griffin (2002) report that about 63% of Hispanics reported as White in C2SS, compared to about 48% in the census, with a correspondingly higher level of "Some other race" reporting in census compared to C2SS. Census 2000 also produced more reports of two or more races than C2SS, especially among Hispanics. Both the census and the ACS collected most cases using self-administered mail

questionnaires with the identical questions ((1) and (3), above) asked in the same order (although the ACS questionnaire used a grid or matrix format). The questions were substantially modified for non-response follow up in both data collections, as discussed above. Although differences in the wording of the nonresponse followup questions may have contributed to reporting differences, interviewers were judged a more likely source of the discrepant results. Census enumerators were temporary hires who received brief training in how to ask questions and carry out their enumeration duties. C2SS was conducted by permanent Census Bureau interviewers who received extensive training for the demographic surveys they work on. Most pertinently, they were trained to probe "Other" race responses to elicit a race category, while census enumerators were not trained to probe such responses (Leslie, Raglin, and Schwede, 2002). Consistent with this explanation (but also with the possible effects of wording variations), Raglin and Leslie (2001), who matched race responses given by the same people in ACS and census and found that Hispanics' race reports were far more consistent when both census and ACS data were collected on mail forms than when they were collected by interviewers in both surveys.

CONCLUSIONS

Serious reporting problems affect race and origin data, especially for Hispanic respondents. Hispanic race reporting is vulnerable to effects of methodological differences between surveys. Reporting of two or more races is also problematic, with results so far showing low reporting reliability.

Census data are the source of the denominators for many race-specific rates based on survey data, and population controls based on census data are used to adjust survey estimates. If the race classification used to produce the denominator is not comparable the race classification on which the numerator is based, then estimates are distorted. Yet, there is a lack of standardization of the race question across surveys and the census, and even across modes within surveys. While some modifications are necessary to administer the question in different modes, the changes sometimes have gone beyond necessary modifications to alter its intent and meaning. Some question wordings do not seem to satisfy the 1997 OMB standards. The introduction of uncontrolled and untested changes in the question on race needs to be reduced to achieve a greater degree of standardization.

The comparisons reported here, and auxiliary investigations of survey discrepancies in race reporting, point to interviewers and survey mode as potentially influencing racial classifications. Careful analyses of existing data and additional field research are needed to evaluate and measure the effects of collection mode upon race reporting. Controlled, experimental studies are needed to investigate and achieve better control over their effects. Again, lack of standardization appears to be a problem which may contribute to the lack of comparable data across surveys.

We believe that the new "one or more" option has increased the complexities of measuring race in important and largely unanticipated ways. Although some versions of the race question appear better than others, there is still much to learn about how to ask about race under the new OMB guidelines. Particular areas where research is needed are:

- Respondent debriefing studies are needed to learn how adequately various versions of the question are communicating the "one or more" option, and to test and refine the wording of the question.
- Some adaptation of the question wording may be necessary in modes that cannot make use of visual aids such as flashcards. Field experiments are needed to test and compare alternative questions

across modes, to ensure they produce comparable data.

- Especially in the absence of a fuller understanding of the effects of question wording and interviewing mode upon race data, we urge great caution in deviating from Census 2000 question wording or race concepts, especially for surveys which require race data comparable to the census and/or which use census race data for their denominators.
- Manipulating question wording, sequencing, and other questionnaire design features as means for achieving response improvements may have inherent limits. When data requirements do not coincide with natural cultural concepts, questionnaire designers may be confronted by trade offs. Increasing adherence to federally mandated categories may be achieved at a cost of increases in nonresponse and larger differences between groups in the reliability of reporting.

REFERENCES

Bates, N., Martin, E., DeMaio, T., and de la Puente, M. (1995) "Questionnaire Effects on Measurements of Race and Hispanic Origin," Journal of Official Statistics 11:435-462.

Baucom, K. (1970) The ABC's of Literacy: Lessons from Linguistics. Hulton Educational Publications.

Bennett, C. (2003) "Exploring the Consistency of Race Reporting in Census 2000 and the Census Quality Survey." Paper presented at the annual meeting of the American Statistical Association, San Francisco, CA.

Bennett, C. and Griffin, D. H. (2002) "Race and Hispanic Origin Data: A Comparison of Results from the Census 2000 Supplementary Survey and Census 2000." Proceedings of Am. Stat. Assoc. (Surv. Res. Methods Sec.)

Bentley, M., Mattingly, T., Hough, C., and Bennett, C. (2003) Census Quality Survey to Evaluate Responses to the Census 2000 Question on Race: An Introduction to the Data. Census 2000 Evaluation B3. Census Bureau.

Davis, D. K., Blair, J., Crawley, E.L., Craig, K. M., Rappoport, M. S. B., Baker, C. A., and Hanson, S.L. (2001) Census 2000 Quality Survey Instrument Pilot Test. Report prepared for the Census Bureau by Development Associates.

Fay, R. E. (1998) VPLX Program Documentation, Vol. 1. U. S. Census Bureau.

Fisher, S., Fricker, S., and Schwartz, L. (2000) Final Report: Results of cognitive testing of the CPS Race and Ethnicity Supplement. Bureau of Labor Statistics, March, 2000.

Gerber, E., de la Puente, M., and Levin, M. (1998) Race, Identity, and New Question Options: Final Report of Cognitive Research on Race and Ethnicity. Census Bureau.

Gould, S. J. (1996) Mismeasure of Man. New York: Norton

Harris, D. R. (2002) "Does It Matter How We Measure? Racial Classification and the Characteristics of Multiracial Youth." In J. Perlmann and M. C. Waters (eds.) The New Race Question: How the Census Counts Multiracial Individuals. New York: Russell Sage Foundation.

Kissam, E., Herrera, E., and Nakamoto, J. M. (1993) Hispanic Response to Census Enumeration Forms and Procedures. Report prepared by Aguirre International for the Census Bureau.

Leslie, T., Raglin, D., and Schwede, L. (2002) "Understanding the Effects of Interviewer Behavior on the Collection of Race Data." Proceedings of the Am. Assoc. for Public Opinion Research, (Survey Res. Methods Sec. of Am. Stat. Assoc.): 2063-2068.

Logan, J. (2002) Hispanic Populations and their Residential Patterns in the Metropolis. Report issued by Lewis Mumford Center for Comparative Urban and Regional Research, University of Albany, NY, May 8.

Los Angeles Times. (2001) "Decline of Latino Groups in Census Has Agencies Angry, Experts Puzzled," August 10, 2001.

Lucas, J. W., Parker, J. D., and Kingston, R. (1999) "Black, White, and Shades of Gray: The Impact of the New Standards for Race and Ethnicity Data Collection on Assessing Race and Health Outcomes in Epidemiologic Research." Paper presented at the annual meeting of the Am. Public Health Assoc., Nov. 8-11, Chicago IL.

Martin, E. (2002) Questionnaire Effects on Reporting of Race and Hispanic Origin: Results of a Replication of the 1990 Mail Short Form in Census 2000. Final report of Census 2000 Alternative Questionnaire Experiment. U. S. Census Bureau.

Martin, E., Gerber, E., and Redline, C. (2003a) Synthesis Report: Census 2000 Alternative Questionnaire Experiment. Census 2000 Alternative Questionnaire Experiment, U.S. Census Bureau, September 26.

Martin, E., Sheppard, D., Bentley, M., and Bennett, C. (2003b) Results of 2003 National Census Test of Race and Hispanic Questions. U. S. Census Bureau, report distributed to the Census Advisory Committees, October 1.

McKenney, N. D., Fernandez, E. W. and Masamura, W. T. (1985) "The quality of the race and Hispanic origin information reported in the 1980 census." Proceedings of the Am. Stat. Assoc. (Survey Res. Methods Sec.): 46-50.

Office of Management and Budget. (1997) "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Part II." Federal Register62(210):58782-90.

Phinney, J.S. and Alipuria, L. L. (1996) "At the Interface of Cultures: Multiethnic/Multiracial High School and College Students." Journal of Social Psychology 36: 139-58.

Raglin, D. and Leslie, T. (2002) "How Consistent is Race Reporting between the Census and the Census 2000 Supplementary Survey?" Proceedings of Am. Stat. Assoc. (Surv. Res. Methods Sec.): 2826-2831.

Rasinski, K. A., Mingay, D., and Bradburn, N. M. (1994) "Do Respondents Really "Mark All That Apply" on Self-Administered Questions?" Public Opinion Quarterly 58:400-408.

Scarr, H. A. (1993) "Review of Federal Measurements of Race and Ethnicity." Census Bureau testimony provided for a hearing of the House Subcommittee on Census, Statistics, and Postal Personnel, April 14, 1993.

Schwartz, L, Fricker, S., and Dixon, J. (2001) Final Report: Results of Behavioral Coding of the CPS Race and Ethnicity Supplement. Bureau of Labor Statistics.

Scott, J. (2001) "Sociologist Offers New Estimates of City Hispanic Census Groups" New York Times, July 6, 2001.

Sheppard, D., Bentley, M., and Woltman, H. (2003) 2003 National Census Test: Analysis of Alternative Race and Ethnicity Questions Based on Panel Comparisons. 2003 National Census Test Analysis. U.S. Census Bureau.

Smiley, R., and Keeley, C. (1997) 1996 Community Census--ICM Behavior Coding Report. 1996 Community Census Results Memorandum No. 23. U. S. Census Bureau.

Tizard and Phoenix, (1995) "The Identity of Mixed Parentage Adolescents" Journal of Child Psychology and Psychiatry, 36(8):1399-1410.

U. S. Census Bureau. (1983) 1980 Census of Population, Volume 1: Characteristics of the Population. Appendix B. Definitions and Explanations of Subject Characteristics, pp. B-3-4.

U. S. Census Bureau. (1996a) Findings on Questions on Race and Hispanic Origin Tested in the 1996 National Content Survey. Population Division Working Paper No. 16.

U. S. Census Bureau. (1996b) Report on the Cognitive Research for the Simplified Enumerator Questionnaire. Unpublished Nov. 18, 1996 report.

U. S. Census Bureau (2001) "Population by Race and Hispanic or Latino Origin for the United States: 1990 and 2000." Census 2000 PHC-T-1.

Xie, Y. and Goyette, K. (1997) "The Racial Identification of Biracial Children with One Asian Parent: Evidence from the 1990 Census." Social Forces 76:547-70.

APPENDIX: Experimental Design for 2003 National Census Test

	4. EXAMPLES	
TREATMENT	Panels without examples	Panels with examples
2000 Short form	Control	
1. Minimal Hispanic origin changes	1	2
2. Drop Some other race	3	4
3. New instructions	5	6
New instructions with SOR category		7

The core of the experiment consists of Panels 1-6, which comprise a 3 X 2 balanced, orthogonal design. TREATMENT has three levels: (1) Wording changes, (2) dropping Some other race, and (3) revised instructions. Each level includes the experimental questionnaire design features incorporated in lower levels, e.g. Panels 3 and 4 include wording changes as well as dropping Some other race. TREATMENT is crossed with EXAMPLES (with two levels, with and without).

Two additional panels were also included. A control (Census 2000) form is included for comparison with Panel 1, to evaluate the effects of the wording changes. Panel 7 includes wording changes, a modified instruction, and retains the Some other race category, to determine if the benefits of a modified instruction apply to a race question with a SOR category.

Each panel was allocated 20,000 sample addresses. The response rates were very similar (about 67 percent) for all eight panels. Thus, none of the experimental treatments significantly affected cooperation.

The control (Census 2000) wording of Hispanic origin was:
Is Person 1 Spanish/Hispanic/Latino? Mark (X) the "No" box if not Spanish/Hispanic/Latino.

No, not Spanish/Hispanic/Latino
Yes, Mexican, Mexican Am., Chicano
Yes, Puerto Rican
Yes, Cuban
Yes, other Spanish/Hispanic/Latino- *Print group*

The panel 1 (revised) wording was:
Is Person 1 of Spanish, Hispanic or Latino origin? Mark (X) "No" if not of Spanish, Hispanic or Latino origin.

No, not of Spanish, Hispanic or Latino origin
Yes, Mexican, Mexican Am., Chicano
Yes, Puerto Rican
Yes, Cuban
Yes, another Spanish, Hispanic or Latino origin- *Print origin*