**Methodological Influences on Comparability
of Race Measurements:
Several Cautionary Examples**

Elizabeth A. Martin* and Eleanor Gerber

Director's Staff*
Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Methodological Influences on Comparability of Race Measurements:
Several Cautionary Examples
Elizabeth Martin and Eleanor Gerber
U. S. Census Bureau

Several recent developments raise concern 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 question on race to produce comparable data in different modes.
- Survey differences in race data appear to be due to question wording and mode differences, and possibly interviewer effects.

We discuss several methodological problems affecting race measurement, including wording and conceptual issues, mode effects, and interviewer effects. Drawing on evidence from qualitative and quantitative research, we comment on wording differences between alternative questions on race. Finally, we suggest research needed to address measurement problems.[1]

RACE AS A SOCIAL CONSTRUCT

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 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

categories[2], and specifies certain aspects of how race and ethnic data are to be collected and reported. Race and ethnicity are regarded as separate constructs (thus, Hispanic people may be of any race). 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).

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. That is, it is possible to report the 'wrong' answer by OMB standards.

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 profound 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.

---

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

[2]The minimum race categories are: *"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 "a person of Cuban, Mexican, Puerto Rican, Cuban, *(sic.)* South or Central American, or other Spanish culture or origin, regardless of race." (OMB, 1997:58789)

Multiracial 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

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. Qualitative research allows some insight into two patterns of approaching the issue.

First, 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 approach to race, rather than a social one. 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. Xie and Goyette (1997) found that Asian-White children's self-identification can be considered "optional." Other research points to the flexibility of self-labels, which 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 there is "clear evidence of variability of multiethnic student's ethnic self-labels in different contexts." Harris (2002) finds that multiracial teenagers offer different responses depending on the presence or absence of their parents during an interview.

Thus, multiracial individuals may offer different accounts of their identification at different times, for complex social and psychological reasons. This may result in poor reliability in multiracial reporting in surveys. Questions that elicit one, stable identifier may be difficult or impossible to construct for persons with such socially fluid identities.

METHODOLOGICAL INFLUENCES ON REPORTING

Question Wording Issues

Two challenges arise in developing a survey question on race that adequately operationalizes the official OMB standards: communicating an appropriate concept of race, and communicating the option of choosing more than one race category. In the course of developing new questions to satisfy the revised OMB standards, a large number of alternative questions (shown in Chart 1) were tested. They illustrate several wording problems and pitfalls.

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. A number of questions in Chart 1 do not meet this test. For example, questions which use the phrase "describe his/her race" (e.g., 5 and 11) are inconsistent with self-identification. "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. This negates the principle of self-identification by making the concept physically, rather than socially, based. Phrases such as "consider him/herself to be" (cf. 1, 3, 6, and 10) or "identify" are preferable because they clearly mark the question as referring to a subjective, rather than physical, set of facts.

Another wording (cf. 7 and 8) is "best indicate ....'s race." While "indicate" is not problematic, the phrase "best indicate" 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 question that asked, "Which of these categories best describes each person's race?" which some respondents interpreted as asking them to choose which race they thought

was the best race (U. S. Census Bureau, 1996b). The follow-up to question 10 ("Which one of these groups...would you say BEST represents your race?") also invites this problematic interpretation. This may contribute to high item nonresponse rates of 23% for White/Black and 13% of White/Asian or Pacific Islander respondents for the question (Lucas et al. 1999). In order to avoid implying invidious comparisons among race groups, we recommend against using "best" or "better" in the question.

Communicating the "one or more" option. For most people, being asked about and reporting their race is very familiar, and this makes it difficult to get them to notice novel features of the question, such as the one or more option. This creates a need for a question that puts some emphasis on the one or more option without leading respondents toward it.

Reading behavior. The way respondents read questions accounts for some of the difficulty of getting them to notice the "one or more" option in mail questionnaires. 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, 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 can be indicated with plural grammatical forms, such as "race or races" (4 and 9) or the plural verb "indicate" in 7 and 8. Such grammatical features are frequently overlooked in reading (Baucom, 1970). They are somewhat more effective in spoken mode, and have been used in interviewer-administered surveys.

Another 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 conducted for the CPS Supplement, which used this probe with question 11, found that interviewers asked the probe in about half of interviews, indicating a failure of standardization (Schwartz, Fricker, and Dixon, 2001). Cognitive testing of question 11 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.

Redundancy. The difficulty of absorbing the option in personal or telephone interviews may be exacerbated because respondents are engaged in competing, simultaneous mental tasks. The tasks of identifying the main subject of the question (race), following an instruction about which person to report for, and listening to a long list of categories may crowd out recognition of the multiple response option. The solution adopted in question 9, which we recommend more generally, was to repeat the multiple response option, giving the respondent more than one chance to absorb it. The option

Chart 1. Illustrative Race Questions from Census Bureau Surveys

---

1. 1990 Census Mail Short Form: "**Race**. Fill ONE circle for the race that the person considers himself/herself to be."

2. 1990 Census Enumerator Questionnaire (for nonresponse follow-up): "What is ...'s race? For example, White, Black, American Indian, Eskimo, Aleut or an Asian or Pacific Islander group such as Chinese, Filipino, Hawaiian, Korean, Vietnamese, Japanese, Asian Indian, Samoan, Guamanian, and so on. *Fill ONE circle for the race that the person considers himself/herself to be.*"

3. Census 2000 Mail Short Form: "**What is this person's race? Mark (X) one or more races** to indicate what this person considers himself/herself to be."

4. Census 2000 Enumerator Questionnaire: "Now choose one or more races for each person. Which race or races does each person consider himself/herself to be?"

5. Accuracy and Coverage Evaluation Survey (CAPI or CATI): "I'm going to read a list of race categories. Please choose one or more categories that best describe [NAME'S] race."

6. American Community Survey Mail Form: "**What is this person's race? Mark (X) one or more races** to indicate what this person considers himself/herself to be."

7. American Community Survey CATI: "I am going to read you a list of race categories. Please choose one or more categories that best indicate (Name's/your) race."

8. American Community Survey CAPI: "Using this list, please choose one or more categories that best indicate (Name's/your) race."

9. Census Quality Survey (telephone survey): "Now I'd like you to tell me what race or races [you consider yourself to be] [he/she considers himself/herself to be]. Please choose one or more of the following 6 race categories:"
"Which of the following Asian groups [are you/is he/she]:"
"Which of the following Native Hawaiian or Other Pacific Islander groups [are you/is he/she]:"
"What other race group is that?"

10. National Health Interview Survey: "What race [do/does] [you/NAME] consider [yourself/ himself/herself] to be? Please select 1 or more of these categories."
[IF MORE THAN ONE]: "Which one of these groups, that is (READ GROUPS) would you say BEST represents [your/NAME's] race?"

11. July 2000 Current Population Survey Race and Ethnicity Supplement: "Please select one or more of the following categories to describe his/her race."

---

might be repeated on a flashcard or in the question. For example, both questions 4 and 9 use two sentences, with each sentence communicating the option in a slightly different way.

Internal Contradiction. Another problem is confusion caused by contradictory elements of the race question. When one part of the question uses a singular concept and another uses a multiple concept, respondents are unsure whether they may

report more than one race. For example, question 3 asks "What is this person's race?" and adds the "one or more" concept in an instruction. In questions 7 and 8, the "one or more" concept is contradicted by the phrase "best indicate" and the reference to the singular "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.

Contradictions may also occur between elements of a race question and the flashcard. The flashcard for question 8 is headed by: "What is this person's race" with no indication that a respondent might choose one or more categories on the card. The phrase "best indicate" in question 8 further reinforces a singular race concept. The flashcard for question 11, with the heading "RACE," also fails to communicate the option.

Distinguishing questions about race and ethnicity. Although the OMB regards 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., Bates, et al., 1995; 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. Reversing the order of the two items to ask Hispanic origin first and adding an instruction ("Please answer BOTH questions 7 and 8") reduces the apparent redundancy and allows Hispanic respondents to report their Hispanic origin before responding to the question on race. The result is reduced Hispanic origin nonresponse and fewer Hispanics reporting as "Some other race" (Bates et al., 1995; U. S. Census Bureau, 1996a).

Racial Classifications and Categories
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 engender emotional reactions and affect respondents' understanding of which, if any, category they belong in. Seemingly equivalent categories may not have the same meaning, and may elicit different reactions from different groups. For example, the term "African American" was an obstacle for Black respondents from the Caribbean or Africa, who did not see themselves as "African American." The inclusion of the term made these respondents hesitate to mark the 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. For example, in 1990 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." The Hispanic origin item also included examples in the 1990 census. Examples were dropped from both questions in Census 2000.

The choice of particular examples can distort reporting. In the 1990 census, "German" was added and "English" was dropped as examples in the ancestry question, and the apparent consequence was more people reporting German ancestry and fewer claiming English ancestry, compared to the 1980 census (Scarr 1993). Examples may enhance recall for example items but inhibit it for nonexample items, as appeared to be the case for ancestry in 1990.

Examples also 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. A questionnaire experiment conducted during Census 2000 confirmed that a question with examples elicited more detailed reports of Hispanic origin (such as "Colombian" or "Salvadoran") than a question without examples, which obtained more generic reports such as "Hispanic," "Latino," or "Spanish" (Martin 2002).

Potential mode effects
 It is difficult to disentangle the effects of mode from self-selection, because experimental research 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. Mode differences in race reporting can result from the effects of mode or from differences in the populations enumerated by each mode, or both.

Questions are sometimes reworded considerably when adapted for different modes in the same survey. For example, compare questions 1 and 2 (used in the mail questionnaire and in nonresponse followup in the 1990 census), questions 3 and 4 (used in the mail questionnaire and in nonresponse followup in Census 2000), and questions 6, 7, and 8 (used in the mail

questionnaire and in CATI and CAPI nonresponse followup in the American Community Survey). Some rewording may be necessary to administer the question in different modes, but many of the changes shown in Chart 1 are substantive changes which affect the meaning of the question.

Communicating the race categories. Because the list 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, as was used with questions 4 and 8. Reading a long list without a flashcard is problematic, as is readily apparent in the version (no. 2) asked by enumerators in the 1990 census: "What is ...'s race? For example, White, Black, American Indian, Eskimo, Aleut or an Asian or Pacific Islander group such as Chinese, Filipino, Hawaiian, Korean, Vietnamese, Japanese, Asian Indian, Samoan, Guamanian, and so on." Attending to this question would impose a large cognitive burden on those conscientious respondents who tried to comprehend and keep the numerous categories in mind before choosing one of them. More likely, respondents broke in with an answer (whether or not it matched a predefined category).

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, as in the CQS (see question 9).

Interviewer Effects
As suggested in the previous sections, 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. One behavior coding study found that interviewers made major changes to the race question (no. 5) in over 40 percent of the interviews that were behavior coded, usually by omitting categories, and skipped it in 7 percent of interviews (Smiley and Keeley, 1997).

EFFECTS OF METHODOLOGICAL DIFFERENCES

For the most part, the effects of different methods for measuring race have not been systematically evaluated (except for the effects of question sequence in mail questionnaires, as discussed above). However, several comparisons among and within surveys suggest the effects of methodological differences may be substantial.

1. 1990-style and Census 2000-style mail short forms
In order to evaluate the net effects of substantive and design changes introduced in Census 2000, data from 1990-style and Census 2000-style mail short forms were compared in an experiment conducted during Census 2000 (Martin 2002). As expected, Table 1 shows more reporting of two or more races in 2000-style forms, which allowed this option, than in 1990-style forms, which asked respondents to "fill ONE circle."

There is a significant difference in reporting as White (alone) by non-Hispanics, probably due to more reporting of two or more races in 2000-style forms. The opposite trend is observed for Hispanics: the fraction of Hispanics reporting as "White" is higher by about 10 percentage points, and the fraction reporting as "Some other race" is lower by about the

Table 1. Race distributions, by mail form type

	Non-Hispanics		Hispanics	
Form type	2000-style	1990-style	2000-style	1990-style
Total	100.0	100.0	100.0	100.0
White	**81.2**	**82.4***	**49.0**	**39.9***
Black	12.3	12.0	2.1	2.3
AIAN	.4	.5	1.5	.7
Asian	4.4	4.3	.6	.9
NHOPI	**.2**	**<.1***	<.1	.2
Some other race	.2	.2	**39.0**	**51.5***
Two or more races	**1.5**	**.5***	**7.8**	**4.6***

*Form difference significant at p<.10. Missing data dropped.

same amount, in the 2000-style forms compared to 1990-style forms, probably due to the effects of item sequence. In addition, more people reported as Native Hawaiian or Other Pacific Islander (NHOPI) in 2000-style forms, perhaps because separating the Asian and NHOPI categories and other design changes helped them find an appropriate category.

These questionnaire effects may confound comparisons of 1990 and 2000 census data. The degree of confounding cannot be inferred from the results in Table 1, which involves relatively small samples, is restricted to mail short forms and does not employ fully edited data. However, we can infer that

the differences in the designs of 1990 and 2000 mail short forms would have resulted in an increase from the 1990 to the 2000 census in Hispanics' reporting of White race, *in the absence of true change in the racial identifications of the population.* The percentage of Hispanics who reported as White (alone) was 51.7 in the 1990 census and 47.9 in Census 2000 (U. S. Census Bureau, 2001). Table 1 implies that the decline in White reporting would have been even larger had the 2000-style questionnaire not increased Hispanics' reporting as White, compared to a 1990-style questionnaire.

2. Accuracy and Coverage Evaluation Survey (A.C.E.) and Census 2000

The A.C.E. was a large reinterview survey conducted shortly after the census to measure, and possibly correct, net undercounts and overcounts that affected the completeness and accuracy of Census 2000 population counts. The survey was intended to provide *independent* and *comparable* information about the numbers and characteristics of household residents on Census Day, 2000. Information obtained in the survey was matched against census enumerations conducted in the households.

Although comparable measures of race are needed for matching and estimation purposes, methodological differences between the Census and the A.C.E. may have produced substantial distributional differences and reporting unreliability. Biemer and Woltman (2001) analyzed race distributions obtained in A.C.E. and the 1998 Census Dress Rehearsal. They found that White and Black race were reported reliably for non-Hispanics. However, for Hispanics, all races were reported with low reliability, and "Two or more races" was not reported reliably by Hispanics or non-Hispanics. Twice as many people reported more than one race in the census (6.8% compared to 3.4% in A.C.E.). Reporting of Some other race was consistently higher in the A.C.E., especially for Hispanics.

Inconsistent race reporting continued into the census itself, according to Farber (2001). Race was inconsistently reported for 3.2% of the census persons who were matched to persons reinterviewed in A. C. E. and whose race and Hispanic origin data were not imputed. The rate of inconsistent reporting exceeded 10% for off-reservation American Indians and Alaska Natives, NHOPI, and Asians.

Several factors likely contributed to inconsistent race reporting between A.C.E. and census.
- Question wordings were not the same. The A.C.E. question (no. 5) used the "best describe" wording, while the Census 2000 mail and enumerator questionnaires (nos. 3 and 4) used the "considers himself to be" wording.
- Interviewing mode was different: Census 2000 was administered by mail (in 64% of households) and personal interviews (36%), while A.C.E. was administered by telephone (in 29% of households) and in person (71%).
- Data were edited differently. In the census, race write-in responses were coded and the data pre-edited in

accordance with Census 2000 procedures. In A.C.E., write-in entries were not coded, and racial classifications were based on the categories marked by the interviewers.

3. American Community Survey (ACS) and Census 2000

A third example is based upon comparisons of race data from Census 2000 and the ACS, which is envisioned as the replacement for the decennial long form in 2010. During Census 2000, the ACS was also conducted in a special national sample (the Census 2000 Supplementary Survey, or C2SS) to establish the operational feasibility of conducting the census and ACS simultaneously, and to assess the quality and comparability of ACS and census data. Both the census and the ACS collected most cases using self-administered mail questionnaires with identical questions 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 seen in Chart 1.

Initial comparisons revealed sizable differences in race distributions, particularly for Hispanics, as shown in Table 2. About 63% of Hispanics reported as White in C2SS, compared to about 48% in the census. There is 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.

Table 2. Race distributions for Hispanics and non-Hispanics, in C2SS and Census 2000 (Household population only)

	Non-Hispanics		Hispanics	
	Census	C2SS	Census	C2SS
Total	100.0	100.0	100.0	100.0
White	**79.3**	**79.6***	**47.9**	**62.9***
Black	**13.5**	**13.2***	**2.0**	**1.6***
AIAN	**.8**	**.8***	**1.2**	**.9***
Asian	**4.2**	**4.3***	**.3**	**.3***
NHOPI	.1	.2	.1	.2
SOR	**.2**	**.2***	**42.2**	**29.4***
Two or more races	**1.9**	**1.7***	**6.3**	**4.8***

*Census and C2SS significantly different at p<.10.
Source: Bennett and Griffin (2002).

Although differences in the nonresponse followup questions (4 vs. 7 and 8) may have contributed to reporting differences, interviewers were thought a more likely source of the discrepant results. Census enumerators are temporary hires who receive brief training in how to ask questions and carry out their enumeration duties. C2SS was conducted by permanent Census Bureau interviewers who receive extensive

training for the demographic surveys they work on. Most pertinently, they are trained to probe "Other" race responses to elicit a race category (or to record race based on observation), while census enumerators are not trained to probe such responses (Leslie, Raglin, and Schwede, 2002). Corroborating evidence is provided by 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.

4. Census Quality Survey (CQS) and Census 2000

In order to provide a "bridge" between old and new methods for obtaining race information, the Census Bureau conducted the CQS, in which old and new questions were asked of the same households. The data can be used to construct statistical adjustments to race distributions obtained using one method to make them more comparable to race distributions obtained using the other.

The survey replicated Census 2000 data collection methods about a year after the census. Both census and CQS (panel A; see Bentley et al. 2003) asked the same "one or more" questions, were conducted using the same methods (mailing out short forms and following up with a personal visit to nonrespondents), interviewed if possible the same respondents, and used the same coding and editing procedures.

People were matched to their census responses to examine the consistency of reporting. Despite the comparability of the methods used in the two surveys, the consistency of reporting two or more 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.

A few months after the CQS interview, an independent telephone reinterview was conducted. Respondents were again asked to report their race, but this time were asked to choose one race (using an altered version of question 9 in chart 1, with a follow-up probe for respondents who persisted in reporting two races). These data were intended to provide a "bridge" between the old and new versions of the race question.

The results suggest problems of unreliability as well as difficulties in bridging from multiple to single race responses. Unreliability is suggested by the fact that 9.4% of respondents who reported two races in the initial CQS interview reported completely different races in the reinterview. This includes 6.9% who reported a race different from either race initially reported, and 2.5% who continued to report more than one race, all of them different from those initially reported. Difficulties obtaining bridging data are indicated by the fact that 16% of respondents would not provide single race answers, even with a probing question. Additionally, the recontact question was not answered by about 13% of respondents. Bennett (2003) finds that the consistency of reporting varies for particular racial combinations.

CONCLUSIONS

Serious reporting problems affect race 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 frequently 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.

Biemer, P. P., and Woltman, H. (2001) "Estimating Reliability and Bias from Reinterviews with Application to the Census Race Question." Paper presented at the Federal Committee on Statistical Methodology Research Conference.

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.

Farber, J. (2001). Accuracy and Coverage Evaluation: Consistency of Post-Stratification Variables. DSSD Census 2000 Procedures and Operations Memo. Series B-10.

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.

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. Census Bureau.

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 Register 62(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.

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.

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