

THE EFFECTS OF SELF AND PROXY RESPONSE STATUS ON THE REPORTING OF RACE AND ETHNICITY¹

Brian A. Harris-Kojetin, Arbitron, and Nancy A. Mathiowetz, University of Maryland
Brian Harris-Kojetin, The Arbitron Company, 9705 Patuxent Woods Dr., Columbia, MD 21046

KEYWORDS: Proxy Reporting

The use of proxy reporters, that is, asking individuals within sampled households to provide information about themselves as well as other members of the household, is an issue that has received attention in both the survey methodological and marketing research literature. For those responsible for the design of a survey, the issue is often framed as a tradeoff among costs, sampling errors, and nonsampling errors, whereby the use of proxy informants can increase the sample size (and hence reduce sampling error) at a lower marginal data collection cost than increasing the number of households. The tradeoff comes at the expense of increased nonsampling error, specifically response error associated with poorer quality reporting for others as compared to the quality that would have been obtained under a rule of all-self response.

Most of the evaluations of the quality of proxy responses as compared to self reports have focused on the reporting of autobiographical information, with some recent investigations examining the convergence of self and proxy reports of attitudes. Empirical investigations have covered a wide variety of topics, ranging from the reporting of labor force participation, health care utilization, and crime victimization, to a variety of attitude questions (for reviews see Moore, 1988; Sudman, Bradburn, and Schwarz, 1996).

Despite all of the empirical investigations concerning the quality of proxy reports and the convergence between self and proxy reports, the literature is silent with respect to the quality of proxy reports for personal attributes or characteristics. Although it is difficult to conceive that static, well-rehearsed characteristics such as gender and age or date of birth may be subject to proxy response effects, other characteristics such as race/ethnicity, occupation, or education may also be vulnerable to proxy response effects to some degree.

The research presented here draws on a recent experiment investigating alternative wording of race and ethnic questions conducted as a supplement to the Current Population Survey (CPS). In this paper, we examine the

consistency in the reporting of racial and ethnic information as a function of the type of report, specifically self or proxy response. Although there was no experimental control with respect to the allocation of self and proxy response status in this investigation, the findings suggest that assumptions concerning the quality of proxy reports for personal characteristics need to be reexamined.

Reporting Race and Ethnicity

In today's society, what is meant when we ask an individual to classify him- or herself or someone else with respect to race or ethnic origin? These concepts may often be treated by survey researchers as objective and fixed characteristics; however, there appear to be many subjective aspects (Smith, 1984) as well as contextual influences on a person's identification (e.g., Bates, Martin, DeMaio, & de la Puente, 1995). In fact, the collection of these data by the Federal Government has repeatedly emphasized the personal, social, and cultural nature of these constructs rather than fixed, determinant characteristics (OMB, 1997, p. 36881).

Although reliable measurement may call for the use of clear, unambiguous, and objective definitions, it is questionable whether these goals are achievable with respect to the measurement of race and ethnicity in the United States. As the population becomes more heterogeneous over time, what was once conceived as a static measure may increasingly be viewed as a more dynamic concept related to an underlying interface between biological definitions, cultural identity, and behavioral norms. Indeed, several theories of racial and ethnic identity focus on how identity (one component of which is racial and ethnic identification) changes over time and contexts (for a review see Phinney, 1990).

How is information about race and ethnicity processed by individuals, either for themselves as well as for others within their family or household? We hypothesize that such information is but one characteristic of an individual that may or may not have been subject to discussion and rehearsal, not unlike information about occupation. As an actor, we may have a well-defined notion of such trait information and our present identity; however, as an

¹ This research was conducted while Brian Harris-Kojetin was at the Bureau of Labor Statistics. The views expressed in this paper are those of the authors and do not reflect the views or opinions of the Bureau of Labor Statistics.

observer we can only draw on the information encoded through observation and/or through discussion with the actor. Thus, it would seem likely that there would be some inconsistency between self- and proxy-reports of race and ethnicity.

Empirical Investigations of Racial and Ethnic Reporting

There has been a variety of research studies addressing the consistency of racial and ethnic identification over time and agreement between different reporters. Specifically, there is some research on the consistency of racial and ethnic reporting by the same person, usually a household respondent, over time. This research has generally shown that consistency in the reporting of race varies depending on the racial group.

For example, McKenney, Bennett, Harrison, and del Pinal (1993) found high consistency in reporting of race for the white, black, and Asian and Pacific Islander racial groups, but considerably less consistency for the American Indian group. Recent research on the reliability of self-reports of race for persons with multiple race backgrounds has also found much lower reliability than for persons with single race/ethnic backgrounds (Cantor, Kerwin, & Schechter, 1997). These individuals may choose to identify with all of their backgrounds or only one in some circumstances, and another in other circumstances.

A related literature has examined context and questionnaire effects on racial identification. Experimental research on different question ordering, more specifically, putting the Hispanic ethnic origin question before the race question, affected response to the race question and racial identification, especially for Hispanics (Bates et al., 1995).

There are also a limited number of studies concerning the consistency in reports of race/ethnicity/ancestry across both time and different reporters. This research has shown that there is inconsistency in racial classification when responses to self-administered are compared to interviewer-administered questions forms (McKenney, Fernandez, and Masumura, 1985). A more comprehensive study by Hahn et al. (1996) yielded several comparisons: between self reports at two points in time, between self and proxy reports at two points in time (where proxy includes information recorded on death certificates), and between an interviewer's initial classification of race compared to either self, proxy, or death certificate information obtained at follow-up. Overall, consistency concerning primary ancestral identification was low regardless of response status.

The Present Study

The purpose of the present investigation was to examine the determinants of inconsistency in reporting of

race and Hispanic ethnicity, specifically the extent to which inconsistency is related to response status. We hypothesize that consistent classification is a function of the experimental panels as well as response status, i.e., self or proxy reports, and the similarity of the essential survey conditions between when the original CPS race and ethnic origin questions were asked and when the supplement questions were asked. We would expect higher rates of consistency of reporting when self-reports are made on both occasions, when the supplement questions are asked close in time to the supplement questions, and when the interviewer is the same.

Methods: Design of the CPS and the Supplement

The CPS is the monthly household labor force survey for the United States conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics. Approximately 50,000 eligible households are sampled each month in a two-stage clustered design. Households selected for the sample are interviewed for 4 consecutive months, are not interviewed 8 months, and then are interviewed again for 4 consecutive months. Furthermore, in any given month, one eighth of the sample is composed of households participating for the first time (month-in-sample 1; MIS 1), one-eighth the second time, etc. Typically, the first and fifth month interviews are conducted as a face-to-face interview; the majority of the remaining interviews are conducted by telephone. Telephone interviews are completed by both field interviewers from their homes (the same interviewers who conducted the face to face interviews) and from centralized telephone interviewing facilities. During the first interview, the interviewer enumerates all members of the household and obtains race and ethnic origin of each household member. In May 1995, a CPS supplement, that is, a set of questions that are asked following the labor force survey, provided the opportunity to evaluate new questions concerning race and ethnicity. This CPS Supplement was one in a series of studies conducted by the Federal government concerning the measurement of race and ethnicity for the Office of Management and Budget's review and revision of Directive 15, Standards for Racial and Ethnic Reporting (see Tucker et al., 1996).

The supplement was designed to address the effect of having a multiracial category among the list of races and) the effect of adding "Hispanic" to the list of racial categories. The supplement was organized into four panels representing a two-by-two experimental design for studying these effects. Each panel was given to one-fourth of the sample, or about 15,000, households. All respondents in a household received the same set of questions; household members 15 years and older were asked to respond for themselves, and parents answered for children too young to answer for themselves. The

panels were defined as:

Panel 1: Separate race and Hispanic origin questions, no multiracial category;

Panel 2: Separate race and Hispanic origin questions, with a multiracial category;

Panel 3: A combined race and Hispanic origin question, no multiracial category;

Panel 4: A combined race and Hispanic origin question, with a multiracial category.

The ethnic origin question asked as part of the household enumeration for CPS is a very different question from the Hispanic origin question asked in the supplement. The intent of the former appears to be to identify each household member's *primary* ethnic origin; the question is used, however, to classify individuals as to whether their origin is Hispanic or not. Neither the question nor the response categories suggest that for those individuals for whom more than one response category is appropriate, preference should be given to those categories which would result in classification as Hispanic. For most individuals, these race and ethnic origin questions had been answered during a previous month's interview; however, for household members who were in the sample for the first time in May, 1995 (1/8 of the sample) this information was obtained during the same interview as the information from the Race and Ethnicity Supplement.

The Effect of Self and Proxy Reporting on the Consistency of Racial and Hispanic Origin Classification

It is important to reiterate that inconsistency in this research is not directly indicative of response error. None of the four panels of the supplement replicate the original CPS race and ethnic origin questions. Without this replication, we have no measure of simple response variance by which to partition the gross measures of inconsistency into components related to simple response variance and response inconsistency related to either a change in the question or change in response status. Rather, we focus our attention on examining to what extent self and proxy reports differ in the rate of consistency between the two sets of race and ethnicity measures.

The use of different race and ethnic questions in the different panels complicates defining a measure that clearly and simply reflects consistency of reporting. As noted earlier, the CPS has two separate questions for race and ethnic origin. The first two panels of the supplement also use separate questions, but in the third and fourth panels respondents were forced to choose between a racial classification or Hispanic origin. Therefore, we chose to define four approaches to the classification of race and Hispanic origin that would allow us to examine

consistency of reporting as completely as possible given these restrictions.

The first measure combines race and Hispanic origin and reflects panels 3 and 4 of the supplement. The other panels of the supplement and the CPS responses were translated into this format by giving precedence to Hispanic origin over race, i.e., all persons identified as Hispanic were classified that way regardless of their racial identification. For the second measure, we excluded all persons who were identified as Hispanic in either the CPS or the supplement and focused solely on consistency of reporting of race. The third and fourth measures reflect the consistency of reporting Hispanic origin. The third measure reflects the consistency of all persons whether they were classified as Hispanic or not Hispanic. However, the fourth measure includes only those persons who were classified as Hispanic in either the CPS or the supplement. This allows us to more closely examine the consistency of Hispanic identification among persons for whom it is most relevant.

For each of the above measures, persons who were identified as multiracial in panels 2 and 4 of the supplement were counted as consistent if any of the races reported in the supplement matched the race reported in the CPS. In addition, for panel 4, persons who were identified as multiracial and indicated they were Hispanic were classified as Hispanic.

Results

Table 1 presents the rate of inconsistency for individuals classified at the two points in time by the various racial/Hispanic origin configurations described above. The race/Hispanic origin (row 1) classification differs for less than 5 percent of the persons included in the sample, with information provided in panel 2 (separate Hispanic question and multiracial category) showing a significantly lower inconsistency rate than the other three panels (all t 's > 4.2 , all p 's $< .01$). The inconsistency rates for the individual components, race and Hispanic origin, are presented in rows 2 and 3. These inconsistency rates ranged from around 2 to 2.5 percent, but showed the same pattern of differences, with panel 2 significantly lower than the other panels (all t 's > 2.4 , all p 's $< .01$). If we limit our analyses to those individuals classified as Hispanic in response to either the CPS enumeration questions or the Race and Ethnicity supplement, the inconsistency rates vary between nearly 20 percent and 25 percent. Panels 3 and 4, those panels that did not include a separate Hispanic origin question, yield significantly higher rates of inconsistency than panels 1 and 2 (t 's > 2.5 , p 's $< .01$), and panel 2 again showed less inconsistency than panel 1, ($t = 2.4$, $p < .01$).

We turn next to the question of interest for the present research, the variation in consistency rates as a function

of the response status. Table 2 examines the same race and Hispanic origin classifications described in Table 1 by self or proxy response status, pooled across all four supplement panels. As noted above, response status is classified into one of three categories: all self response, all proxy response, and a mix of self and proxy responses. While it is clear that the first category, all self response, by definition, indicates that the same respondent reported race and ethnicity at the time of the original CPS interview and the supplement, and that the last category, mix of self and proxy response status indicates that a different individual reported the information for the CPS interview and the supplement, we are unable to determine (definitively) whether those responses classified as “all proxy responses” were obtained from the same respondent or from different respondents.

Regardless of which classification configuration we examine, the inconsistency rate is lowest for those individuals who reported for themselves for both the original CPS and the supplement (all t 's > 3.2 , all p 's $< .01$). Furthermore, the highest rate of inconsistency is among those persons for whom information was obtained by proxy for both sets of questions on three of the four measures (rows 1-3; (all t 's > 3.0 , all p 's $< .01$). With respect to Hispanic identification, only among those ever identified as Hispanic (row 4), the inconsistency rate across responses obtained all by proxy or by a mix of self and proxy response did not differ ($t = .99$, ns).

The findings presented in Tables 1 and 2 suggest that inconsistency is a function of both the questions included in the supplement as well as the response status of the individual of interest. However, as noted above, we also hypothesize that consistency in the classification of the race and ethnic origin of an individual may also be a function of the time between the original interview and the administration of the supplement and the consistency of the mode of data collection and the use of the same interviewer to conduct the interview.

We examined several multivariate logistic models corresponding to the four race and Hispanic origin consistency measures presented in Tables 1 and 2. Regardless of the way in which we define inconsistency in the reporting of race or Hispanic origin, the findings across all four measures indicate significantly higher rates of inconsistency for reports obtained by proxy or a mix of self and proxy reports as compared to reports obtained by self-response only. These findings are robust even when controlling for the questions used in the supplement, the month-in-sample, the use of the same interviewer, as well as the specific household member's race, the relationship between the respondent and the household member of interest, and various characteristics of the respondent. The effects of these other factors are described below.

Among those individuals ever identified as Hispanic, the use of a separate Hispanic question reduces the rate of inconsistency across the two sets of reports. The use of a separate Hispanic question comes closest to replicating the original set of questions, albeit the original CPS questions focus on *ethnic* origin as opposed to Hispanic origin. Therefore, it is not surprising that in those supplemental panels that employed a separate Hispanic question, the rate of inconsistent reporting drops. In addition, there is some evidence that consistency rates were higher among individuals for whom both a separate Hispanic question and a multiracial response option was offered. Once again, this seems intuitive, given that for those individuals for whom multiracial category was selected, responses were considered consistent if *any* of the racial categories enumerated in the supplement matched the report offered in the initial CPS interview.

When the supplement questions on race and Hispanic origin were asked during the same interview as the original CPS race and ethnic origin questions, inconsistency rates for measures of race tended to be lower than when the two data points were separated by time. In contrast, with respect to Hispanic classification, inconsistency rates were *higher* when the two sets of questions were asked at the same time. Interviews conducted completely by field interviewers (as compared to a mix of field and centralized telephone interviewers) yielded more consistent reporting of race and Hispanic origin, although this finding is not statistically significant across all of the measures of race and Hispanic origin that we examined.

Discussion

The primary objective of the research presented here was to examine the extent to which a supposedly static characteristic such as race or Hispanic origin may be subject to measurement effect related to response status, that is, self reports as compared to reports obtained by proxy. Although the issue of response status is not new to the literature on measurement errors, that literature has focussed almost extensively on the reporting of autobiographic event information and only recently on the reporting of attitudes. With respect to effects of response status on the reporting of trait information, the literature is silent.

The May, 1995 Race and Ethnicity Supplement to the Current Population Survey was designed to test the effects of alternative sets of questions and response categories on the classification of both race and Hispanic origin for the U.S. population. The design of the study was intended to provide information as to the extent to which the classification of individuals with respect to race and Hispanic origin shifts as a function of a separate question on Hispanic origin, the use of a multiracial

response option, and the joint effect of the two experimental features.

Given that the CPS collects race and ethnic origin as part of the enumeration of individuals within the household, the administration of the supplement provides a unique opportunity to not only examine the marginal distributions resulting from the different supplement panels, but also the opportunity to examine the characteristics associated with shifts in the classification at the individual level. It is within this context that we examine response status as one of the characteristics associated with response variance over time.

The original experiment was not designed to address issues of self and proxy response status and thus, this study suffers from the non-randomization of individuals to self and proxy response status. To the extent that individuals for whom data are collected entirely or in part by proxy differ (with respect to race and Hispanic origin) differ from those for whom data are collected entirely by self-response, the findings presented here are confounded. Others (e.g., Moore, 1988; Mathiowetz and Groves, 1985) have addressed the potential confounding between reports of health, crime and employment and response status; similarly, we find evidence that suggests a confounding between race and Hispanic origin and response status.

Our findings suggest that inconsistency in the reporting of race and Hispanic origin is a function of response status, with higher rates of inconsistency among household members for whom the information was collected entirely by proxy report or by a mix of self and proxy report as compared to all self response. These findings were robust regardless of the measure of race or Hispanic origin used and, in general, persisted in models in which characteristics of the target individual as well as the respondent were included.

The research reported here addresses but one aspect of measurement error with respect to the measurement of race and ethnicity, response variance, as a function of response status. The research is uninformative with respect to levels of bias associated with response status. However, to the extent that the reduction of both response bias and variance are goals of standardized measurement, these findings suggest that measures of individual characteristics or traits may be subject to varying response effects as a function of who is reporting for whom. However, the issue warrants further investigation with a design that would ideally randomize individuals to response status and would permit comparison between both self and proxy reports for the same individual.

References

- Bates, N., Martin, E. A., DeMaio, T. J., & de la Puente, J. (1995). Questionnaire effects on measurements of race and Spanish origin. *Journal of Official Statistics, 11*, 433-459.
- Cantor, D., Kerwin, J., and Schechter, S. (1997). Reliability and consistency of self-reports of race among young multiracial mothers. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Norfolk, VA.
- Hahn, R. A., Truman, B. I., & Barker, N. D. (1996). Identifying ancestry: The reliability of ancestral identification in the United States by self, proxy, interviewer, and funeral director. *Epidemiology, 7*, 75-80.
- Mathiowetz, N. & Groves, R. (1985). The effects of respondent rules on health survey reports. *American Journal of Public Health, 75*, 639-644.
- McKenney, N., Bennett, C., Harrison, R., and del Pinal, J. (1993). Evaluating racial and ethnic reporting in the 1990 Census. *Proceedings of the American Statistical Association Section on Survey Research Methods, 66-74*.
- Moore, J. C. (1988). Self/proxy response status and survey response quality: A review of the literature. *Journal of Official Statistics, 4*, 155-172.
- Office of Management and Budget (1997). Recommendations from the Interagency Committee for the Review of Racial and Ethnic Standards to the Office of Management and Budget concerning changes to the Standards for the Classification of Federal data on Race and Ethnicity; Notice. *Federal Register, 62 No. 131*, 36874-36946.
- Phinney, J. S. (1990). Ethnic Identity in adolescents and adults: Review of Research. *Psychological Bulletin, 108*, 499-514.
- Smith, T. W. (1984). The subjectivity of ethnicity. In C. F. Turner and E. Martin (Eds.) *Surveying subjective phenomena*. New York: Russell Sage Foundation.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Tucker, C., McKay, R., Kojetin, B., Harrison, R., de la Puente, M., Stinson, L., & Robison, E. (1996). *Testing Methods of Collecting Racial and Ethnic Information: Results of the Current Population Survey Supplement on Race and Ethnicity*. BLS Statistical Notes #40. Washington, DC: Bureau of Labor Statistics.

Table 1. Percent Inconsistent Responses for Racial and Hispanic Origin Identification Questions between Initial CPS Interview and Supplement by Panel and Various Racial/Hispanic Origin Configurations

Racial/Hispanic Origin Identification	Supplement Panel			
	Panel 1	Panel 2	Panel 3	Panel 4
Combined Racial/Hispanic Origin (n=124,534)	4.5% (.102)	3.9% (.091)	4.7% (.108)	4.7% (.132)
Racial Identification Only (n=113,160)	2.5% (.079)	2.2% (.071)	2.6% (.080)	2.5% (.088)
Hispanic Origin Identification Only (n=124,534)	2.3% (.082)	2.0% (.065)	2.3% (.080)	2.5% (.111)
Hispanic Origin Identification among Those Identified as Hispanic in at Least One Interview (n=11,374)	20.5% (.670)	18.4% (.562)	24.5% (.773)	23.6% (.902)

Source: CPS 1995 May Supplement

NOTE: Panel 1: Separate race and Hispanic-origin questions; no multiracial category
 Panel 2: Separate race and Hispanic-origin questions with a multiracial category
 Panel 3: Combined race and Hispanic-origin question; no multiracial category
 Panel 4: Combined race and Hispanic-origin question with a multiracial category

Table 2. Percent Consistent for Racial and Hispanic Origin Identification Questions between Initial CPS Interview and Supplement by Self/Proxy Response Status and Various Racial/Hispanic Origin Configurations (Across all Four Panels)

Racial/Hispanic Origin Identification	Supplement Panel		
	All Self Response	All Proxy Response	Mix of Self and Proxy Response
Combined Racial/Hispanic Origin (n=124,534)	2.3% (.047)	5.2% (.080)	4.6% (.140)
Racial Identification Only (n=113,160)	1.9% (.035)	2.8% (.063)	2.4% (.102)
Hispanic Origin Identification Only (n=124,534)	1.6% (.038)	2.7% (.055)	2.4% (.107)
Hispanic Origin Identification among Those Identified as Hispanic in at Least One Interview (n=11,374)	20.2% (.453)	22.1% (.377)	22.9% (.832)

Source: CPS May Supplement, 1995

Note: All Self-response occurred when the same person reporting for themselves to the race and origin questions on the basic CPS in the first month-in-sample interview and also reported for themselves on the May 1995 Supplement. A mix of self and proxy response occurred when only the initial CPS race and ethnic information was obtained by self-report or only the supplement information was obtained by self-report. The all proxy-response indicates that proxy-reporters gave both the initial CPS and the supplement race and ethnic information. However, it is not known whether it was the same proxy reporter on both occasions or different proxy reporters