

THIRD-PERSON REPORTING OF HISPANIC ORIGIN AND RACE IN A
GROUP QUARTERS/ESTABLISHMENT CENSUS

by

Laurie Schwede, Ph.D.
Center for Survey Methods Research
Statistical Research Division
U. S. Bureau of the Census

Paper presented at the Annual Meetings of the
American Association for Public Opinion Research
Norfolk, Virginia

May 17, 1997

To be published in the American Statistical Association
1997 Proceedings of the Survey Research Methods Section

This paper reports the results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

THIRD-PERSON REPORTING OF HISPANIC ORIGIN AND RACE IN A
GROUP QUARTERS/ESTABLISHMENT CENSUS

Laurie Schwede, U.S. Bureau of the Census¹
Statistical Research Division, 3133/4, U.S. Census Bureau, Washington, D.C. 20233-9150

KEYWORDS: Race, Hispanic Origin, Group Quarters, Establishment Survey, Third-person Reporting, Pretests

INTRODUCTION

Hispanic origin and race are two of the most problematic data elements in censuses and surveys. These data are needed for monitoring and enforcing civil rights and allocating federal funding. To improve cross-survey comparability, the Office of Management and Budget issued Directive 15 in 1977 (OMB 1977) to standardize the minimum basic categories in federal data collections. These include four mandated categories for race--white, black, Asian/Pacific Islander, and American Indian/Alaskan Native--and one category for Hispanic origin (sometimes referred to as "ethnicity"), which can be collected in either separate or combined race/origin questions, as shown in Table 1.

Table 1: OMB Directive 15 Basic Categories and Rules for Race and Hispanic Origin (1977)

<p><u>Option 1: Separate Race and Origin Questions</u></p> <p><u>Race Categories</u></p> <p>American Indian or Alaskan Native Asian/Pacific Islander Black White</p> <p><u>Origin Categories</u></p> <p>Of Hispanic Origin Not of Hispanic Origin</p>	<p><u>Option 2: Combined Race and Origin Question</u></p> <p><u>Race/Origin Categories</u></p> <p>American Indian or Alaskan Native Asian/Pacific Islander Black, not of Hispanic Origin White, not of Hispanic Origin Hispanic</p> <p style="text-align: center;"><u>Special Rule</u></p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

"When race and Hispanic origin are collected separately, the number of White and Black Hispanics must be identifiable and capable of being reported in that category."

Recognizing that these categories have become less useful over time as the nation has diversified through increases in immigration and intermarriages (OMB 1994, Evinger 1995), OMB began a three-year public review of Directive 15 and requested research to aid in deciding in the fall of 1997 whether and how Directive 15 should be revised. In response, a large amount of research has been done on Hispanic origin and race reporting in the last few years, almost all of it in the context of household surveys where one householder reports origin and race data on a few coresidents, most often well-known to him/her. Very little research has been done in establishment surveys where one respondent, a third-person reporter, may report origin and race data on many persons, some or all of whom may not be known personally by him/her. The mix of factors affecting the accuracy and completeness of Hispanic origin and race data in establishment surveys will differ from those influencing household surveys.

The purpose of this paper is to identify factors that may affect the quality of race and origin data provided by third-person reporters in juvenile facilities. I describe the project background and research methodology, then summarize literature on third-person reporting of origin and race in non-household data collections. I identify and discuss factors that may affect the quality of Hispanic origin and race data in this establishment census, recommend the use of a combined race/origin question for this census, and suggest new research.

BACKGROUND AND METHODOLOGY

Our interest in the quality of race and origin data developed from exploratory interviews conducted in the early stage of redesigning the Children in Custody Census for the Office of Juvenile Justice and Delinquency Prevention (OJJDP) in the Department of Justice. The self-administered questionnaire for this census is mailed to roughly 3600 facilities nationwide and collects data on characteristics of the facilities and the juveniles housed in them. These public and private facilities vary greatly in population size, from 3 to more than 1200 juveniles. The census frame includes a wide diversity of types, ranging from secure institutions such as detention centers and training schools to more open facilities, such as group homes, shelters, and wilderness camps.

From in-person, exploratory interviews with respondents who complete the census forms for these

types of facilities, we identified several potential problems with the race and origin questions: lack of standardized methods among facilities for collecting, recording, and maintaining race information; inconsistencies in degrees of fit between race/origin codes used by the facilities and those mandated by Directive 15; and variations in the extent to which respondents could distinguish white and black Hispanics.

These early findings suggested the existence of factors influencing third-person reporting of race and origin in our facility census that differed from those affecting household surveys. Because race and origin data from this census are used by OJJDP staff in annual reports to Congress on juveniles in facilities, our sponsor wanted to identify problems and improve the quality of these questions. With OMB approval to experiment with race and origin questions, we incorporated race research into our questionnaire pretesting cycle.

We developed and tested a questionnaire with experimental race and Hispanic origin questions and supplemental retrospective questions in two rounds of cognitive interviews. This paper is based on data collected with retrospective questions on 1) facility methods used to obtain race and origin data, 2) race categories used by respondents, and 3) the degree of fit between facility and OMB categories. The data used here come from 40 exploratory interviews and 32 cognitive interviews in 18 states and Washington, D.C.²

REVIEW OF THE LITERATURE

To date, I have not found any previous study on the quality of race reporting in residential establishment surveys: the types of group quarter facilities included in the decennial census. There are some studies on this topic in non-residential establishments or with administrative records. A mailout survey of public schools designed by the U.S. Department of Education National Center for Education Statistics (1996) and conducted by Westat found that about three-quarters of the schools collected self-reported race from the parents, the most reliable method, but that most of the other one-quarter determined race by observation only. They also found that about 73% of schools used just the five standard federal categories, while the rest had extra nonstandard categories of "other" or "multiracial."

Research on this topic has been done in the field of public health. Hahn, for example, has done a series of papers on the validity and reliability of race, ethnicity, and ancestry data in current public health studies and calls for improvement in these concepts (1992). He and Stroup (1994) note that the lack of information on whether doctors determine race by patient self-report or by clinician observation raises questions about the

validity of race data in a national public health surveillance system. Hahn et al. (1996) also compared ancestry data given by self, proxy, interviewer, and funeral director (on death certificates) and concluded that the low reliability of ancestry measures over time and across observers complicates analysis, especially for those neither white nor black. In comparing race on birth and death certificates for infants, he found inconsistencies in reporting were almost 9 times higher for Hispanics than for whites or blacks (1992).

FACTORS THAT MAY AFFECT THE QUALITY OF FACILITY RACE/ORIGIN DATA

In household surveys, the aim is to collect self-reported race and origin, but in practice, this is not always achieved. One household respondent provides a self-report for himself/herself, but gives proxy reports on others in the household. We assume this household proxy reporter gives accurate reports from memory, because the others are most often well-known by him/her.

In the Children in Custody Census, all respondents are third-person proxy reporters on the juveniles in their facilities. The first factor that may affect the quality of race and origin data is that facility respondents vary in their ability to give accurate reports from memory of how resident juveniles would self-report race and origin. Some respondents in small facilities know their juveniles well and can give self-reports from memory. Others, especially those who report on many juveniles, may have little or no personal contact with the children, relying on records for race and origin data. Whether or not they report from memory, all but one of the 72 respondents in both exploratory and cognitive interviews collected race data on juveniles. These data are collected at or near the time of a juvenile's admission to the facility.

The second factor that may affect data quality is wide variation in methods used to collect race and origin data for the facility records. We found three basic methods of collecting facility record data: 1) self-reports of race by the juvenile or parent, 2) intake worker's ascription of race based on observation and/or last name, and 3) reliance on incoming administrative records prepared by persons in other locations or agencies. Table 2 shows the reported methods used in the 32 juvenile facilities included in our cognitive interview research. Eleven respondents obtained race and origin from self-reports by the juveniles or their parents. Fourteen relied mostly on incoming administrative records from police, probation officers, courts, referral agencies, other parts of their own agencies, or Vital Statistics Departments. Of these fourteen, ten relied solely on records, while the other four sometimes asked for self-reports as checks on

on or in disagreements with the incoming administrative record data. Four others reported using both incoming records and intake worker observation, with one of these using self-reports only when needed. One other respondent used observation, with self-report requested only if necessary. Finally, two mentioned all three methods, but were vague about the primary method used.

Table 2: Methods Used in Facilities to Collect Race and Origin Data for Facility Records

<u>Method</u>	<u>N of Facilities</u>
Self-report requested from juvenile or parent	11
Use of incoming administrative records:	14
with occasional self-reports	4
self-reports not mentioned	10
Use of incoming administrative records and intake workers' observation:	4
with self-reports, if necessary	1
self-reports not mentioned	3
Intake worker observation, with self-report only if necessary	1
No consistent answer on methods	2
Total N of Facilities	32

Of the three main methods, self-reporting would produce the most valid and reliable data and is the method preferred by the Office of Management and Budget. Data collected by observation would be subject to inter-observer differences, introducing error both within the facility and among facilities. Observation alone would also be likely to underreport biracial juveniles and those of races rarely seen by personnel in their geographical area.

The third method--use of administrative records prepared elsewhere--raises the most concerns about validity and reliability, since we have no knowledge of how and when these data were originally collected. One respondent just copied into his records whatever the police officers recorded as the juveniles' race, but reported that problems sometimes arose when the police categories did not match those he used. He was not sure how the police obtained the data. Another said that her facility used incoming records, including birth certificates, rather than asking the juvenile or parent about the sensitive topic of race. We have no sense of the

validity of birth certificates issued years ago. We hypothesize that the further removed one is from using self-identified race, the greater the chance of error and misclassification of race and Hispanic origin.

The collection methods reported by our respondents seemed clear-cut, but variations in methods used may occur within facilities over time and among intake workers. It is clear that variation in methods used to collect race data for facility records is a factor with potentially large effects on the validity and reliability of race and origin data within and among facilities.

A third potential factor is the lack of standardized wording for requesting race and origin data. Intake forms may just have the key word "race" and either a set of race categories or just a blank line for an open-ended response.

A fourth factor that may affect the quality of race data is the degree of fit between the federal categories and the facility categories. It should not be assumed that facilities use the federal categories for their own record keeping. Facility race categories may just evolve as new cases come in. We found a lack of standardization in race categories used in the facilities' internal records. One facility used 3 categories while another used 127! Respondents from four small facilities did not keep aggregate race data continuously; when aggregate data were needed for reports or surveys, they would do *ad hoc* counts from memory or from a review of the individual paper files. Most respondents who did keep continuous aggregate data used facility categories similar to those of OMB, but sometimes with different labels that may not completely overlap with the federal categories (Table 3). Some facilities did not have separate response categories for Asian/Pacific Islanders or American Indians. These respondents would include

Table 3: Variations in Federal and Facility Race and Origin Labels

<u>Federal Labels</u>	<u>Facility Labels</u>
American Indian/ Alaskan Native	North American Indian Indian Native Indian Native American
Asian/Pacific Islander	Asian Oriental
Black	African American
Hispanic	Latino Spanish American Mexican National Mexican American

the rare incoming juveniles of these races into an "other" category.³ Other facilities in border states had special categories to separate legal from illegal Mexican immigrants, a distinction not made with the Directive 15 categories.

The fifth factor that may affect how respondents answer Hispanic origin and race questions is the conceptual and operational definition of "Hispanic" as a race, rather than as a separate concept of culture or origin, as defined by OMB. The great majority of respondents told us that "Hispanic" was a race category in their records, along with white, black, and their other categories. We were not surprised to find that just 4 of the 32 cognitive interview respondents kept records disaggregating black and white Hispanics. In the earlier exploratory interviews, not one of the 40 respondents kept records separating white from black Hispanics. This clearly suggests that a combined race/origin question should be used in this establishment census.⁴

This conceptual view of many of our third-person respondents of "Hispanic" as a race had important effects on the patterns of responses to our test origin and race questions (Schwede 1997). This conception of "Hispanic" as a race, rather than as a separate concept such as ethnicity, has been identified as a factor affecting race/ethnicity responses in research on household censuses and surveys, in some studies by Hispanic respondents (Bates et al. 1995, Kissam et al. 1993, Elias-Olivares and Farr 1990, Martin et al. 1990, McKenney et al. 1988) and in others, by both Hispanic and non-Hispanic respondents (Gerber, de la Puente, and Levin 1997, McKay et al. 1996, and Gerber and de la Puente 1996). It is likely that this conception of "Hispanic" as a race affects the quality of race/origin reporting in establishment censuses and surveys of other types of residential facilities but this remains an hypothesis, given the apparent lack of any previous study on race reporting in a residential establishment survey.

Table 4: Extra Race/Origin Facility Categories Not Permitted by Directive 15 (1977) for Federal Data Collection Reporting

<u>Category</u>	<u>N of Facilities</u>
Other	19
Biracial, multiracial	10
Unknown	7

In addition to keeping main categories similar to those of the Office of Management and Budget, quite

a few keep data in extra categories, such as "other," "biracial," and "unknown" (Table 4) that vary in the extent of being foldable into the five basic categories, as OMB requires.⁵ Nineteen of the cognitive interview respondents included an "other" category in their facility race codes. These "other" categories covered persons and situations such as: less frequently encountered minorities (Asian/Pacific Islanders and/or American Indians/Alaskan Natives in some areas), refusals, and more rarely, juveniles of unknown race. Sometimes (but not always) these "other" categories used by facilities also included juveniles who could be identified in the general category "biracial," without the specific mix of races.

Ten of the 32 respondents reported a separate "biracial" or "multiracial" response category in their records. At least five of these said that the label, "biracial," is sufficient for their records, but they could disaggregate these to specify the race mixes by checking the individual paper files, if necessary. Other respondents, including one reporting on all of the several thousand juveniles in public facilities in his state, could not disaggregate the biracial juveniles. The differential ability of respondents to give specific mixes of biracial juveniles would introduce bias.

The decision to use a biracial category in a facility may be made either at the facility or the state level. At least seven states now have state-mandated "biracial" categories: Florida, Georgia, Illinois, Indiana, Michigan, Ohio, and North Carolina (*Emerge Magazine*, December/January 1996, p. 51). Respondents in facilities using "biracial" or "other" categories may have trouble translating juveniles in those categories in their files into the federally-mandated codes, especially if an "other, specify" line is not given on the federal form.

Additionally, seven respondents used the category of "unknown," mainly for refusals. One pointed out that race might be coded as "unknown" for some adopted children. Absent a federal category for "unknown," these respondents have no way of putting juveniles of unknown race into the federal categories.

The sixth factor that may affect the accuracy and completeness of race reporting is variations in the mode of record-keeping. In the cognitive interviews, those who relied on individual paper case files were more likely than those with computer records to decline to go through the files to find individual-level data during the in-person interviews. Some just estimated data, while others refused to complete the form. Some using electronic records had trouble if they had narrative case histories or had to keep switching among files and screens for each juvenile.

Another implication of record-keeping mode is whether or not multiple race codes can be entered. In individual paper case files, any number of races can be

listed for each child. However, with electronic files, it is possible that only one fixed race code is allowed. At least 8 of the 32 facility systems allowed just one fixed race code. If mixed race juveniles can only be entered into computer files with one race, there would be no way to reidentify them by computer later, should some survey ask for numbers and specific mixes of biracial children.

SUMMARY AND RECOMMENDATIONS

The results of exploratory and retrospective interviews have been used to identify six factors that may affect the validity and reliability of race and origin data provided by third-person reporters in juvenile facilities:

- 1) the extent to which respondents can accurately report, from memory, juveniles' self-reported race and origin;
- 2) the methods by which the facility obtains race and origin data on juveniles;
- 3) the extent to which race and origin question wordings are standardized within and among facilities;
- 4) the degree of consistency among categories used by the facility, the state and the federal government;
- 5) the variation in conceptual and operational definitions of "Hispanic" as a race, not as a separate concept of origin or ethnicity; and
- 6) the paper or electronic mode of keeping individual files.

The next step in this research would be to design a survey to test the effects of these factors by comparing the consistency of reported race and origin by third-person respondents and by self-reporting juveniles. Such a study would be useful to the sponsor in assessing the reliability, validity, and completeness of data used in its annual report to Congress on juveniles in facilities. It would be useful to OMB in obtaining data on the validity and reliability of its categories in a non-household census. And finally, such a study would be a starting point for evaluating the quality and completeness of race and origin data collected in group quarter censuses and surveys, since to my knowledge, no such study has ever been done within residential facilities. Improvements in the quality and completeness of data from group quarters might lead to coverage improvements for undercounted ethnic populations in the decennial census as well.

NOTES

1. The author is the manager of the Children in Custody Questionnaire Redesign Project. Laurie Moyer joined in

the first round of question development, testing, and analysis. Catherine Gallagher joined in the next round. The author thanks Eleanor Gerber, Theresa DeMaio, Elizabeth Martin, Laurie Moyer, Joseph Moone, and Martin Wulfe for reviewing earlier versions of this paper.

2. Complete descriptions of the methodology and results of the exploratory and cognitive interviews are found in Schwede and Ott (1995), Schwede and Moyer (1996) and Schwede and Gallagher (1996). Description and analysis of the experimental separate and combined race and origin questions tested in this research are found in Schwede 1997. Exploratory interviews were conducted in Colorado, California, Florida, Maryland, Massachusetts, Michigan, New York, Pennsylvania, Texas, Washington, and Washington, D.C. Cognitive interviews were conducted in Arizona, California, Georgia, Illinois, Indiana, Maryland, Minnesota, Missouri, New York, Ohio, Virginia, and Washington, D.C. The performance of an experimental combined race/origin question in a later split-panel test is found in Ellis and Schwede 1997 (elsewhere in the *Proceedings*).

3. The lack of separate race categories for Asians and American Indians tended to occur in facilities in regions of the country with very low proportions of persons in these racial groups. American Indians and Asians are concentrated in the west and are thus underrepresented in the midwest and northeast, according to 1990 Census results (Harrison and Bennett 1995).

4. Both this paper and the final version of the new "Children in Residential Placement Census" questionnaire had to be finalized between the July 9, 1997 publication of the Interagency Committee for the Review of the Racial and Ethnic Standards recommendations for revising the Directive 15 rules and categories (Office of Management and Budget 1997) and the mid-October, 1997 scheduled date for OMB to announce its final decisions on this matter. The Interagency Committee had recommended that separate race and Hispanic origin questions should be used when self-identification was the method utilized for determining race and ethnicity, but a combined question could be used "when self-identification is not feasible or appropriate" (OMB 1997: 123). Based on the results from this research in juvenile facilities (Schwede 1997), the Office of Management and Budget approved the use of a combined race/Hispanic origin question and an "other" category in the new questionnaire for the October, 1997 census.

5. Additional, more detailed categories may be collected, if they can be aggregated into the 5 categories in reports.

REFERENCES CITED

- Bates, N., E. Martin, T. DeMaio, and M. de la Puente. 1995. "Questionnaire Effects on Measurements of Race and Spanish Origin." in *Journal of Official Statistics*, Volume 11, No. 4, pp. 433-459.
- Elias-Olivares, L. and M. Farr. 1991. "Sociolinguistic Analysis of Mexican-American Patterns of Non-response to Census Questionnaires." Report submitted to the Census Bureau under Joint Statistical Agreement 88-25, Ethnographic Exploratory Research Report # 16.
- Ellis, Y. and L. Schwede. 1997. Evaluation of the Redesign Questionnaire for the Children in Custody Census. *1997 Proceedings of the Section on Survey Research Methods*. Alexandria, Virginia: American Statistical Association.
- Evinger, S. 1995. "How Shall We Measure Our Nation's Diversity?" In *Chance*, Winter, 7-14.
- Gerber, E. and M. de la Puente. 1996. "The Development and Cognitive Testing of Race and Ethnic Origin Questions for the Year 2000 Decennial Census." *1996 Annual Research Conference Proceedings*. Bureau of the Census.
- Gerber, E., M. de la Puente, and M. Levin. 1997. "Race, Identity, and New Question Options: Final Report of Cognitive Research on Race and Ethnicity." Available from the Center for Survey Methods Research, Statistical Research Division, U.S. Bureau of the Census.
- Hahn, R. 1992. "The State of Federal Health Statistics on Racial and Ethnic Groups," in *Journal of the American Medical Association* Volume 267, No. 2. January 8, 1992, pp. 268-271.
- Hahn, R., J. Mulinare, and S. Teutsch. 1992. "Inconsistencies in Coding of Race and Ethnicity Between Birth and Death in U.S. Infants," in *Journal of the American Medical Association*, Vol 267, No. 2, pp. 259-263.
- Hahn, R. and D. Stroup. 1994. "Race and Ethnicity in Public Health Surveillance: Criteria for the Scientific Use of Social Categories," in *Public Health Reports*, Volume 109, No. 1. January-February, pp. 7-14.
- Hahn, R., B. I. Truman, and N. Barker. 1996. "Identifying Ancestry: The Reliability of Ancestral Identification in the United States by Self, Proxy, Interviewer, and Funeral Director," in *Epidemiology* January 1996, Volume 7, No. 1, pp.75-80.
- Harrison, R. and C. Bennett. 1995. "Racial and Ethnic Diversity." Chapter 4 in *State of the Union, America in the 1990s, Vol. 2, Social Trends*. 1990 Monograph Series, Ed. R. Farley. New York: Russell Sage Foundation.
- Kissam E., E. Herrera, and J. Nakamoto. 1993. "Hispanic Response to Census Enumeration Forms and Procedures." Report submitted to the U.S. Census Bureau under Contract No. 50-YABC-2-66027, Task Order No. 46-YABC-2-0001, March.
- Martin, E., T. DeMaio, T. and P. Campanelli. 1990. "Context Effects for Census Measures of Race and Hispanic Origin," in *Public Opinion Quarterly* 54: 551-566.
- McKay, R, L. Stinson, M. de la Puente, and B. Kojetin. 1996. "Interpreting the Findings of the Statistical Analysis of the CPS Supplement on Race and Ethnicity." *1996 Annual Research Conference Proceedings*. Bureau of the Census.
- McKenney, N., A. Cresce, and P. Johnson. 1988. Development of the Race and Ethnic Items for the 1990 Census. Paper presented at the Population Association of America Annual Meeting in New Orleans, Louisiana, April, 1988.
- Office of Management and Budget. 1997. "Recommendations from the Interagency Committee for the Review of the 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. July 9, 1997 Federal Register. Office of Management and Budget. 1994. "OMB Announces Review of Federal Racial and Ethnic Categories and Related Public Hearings." Washington, D.C. Federal Register, June 9, 1994.
- Office of Management and Budget. 1977. Statistical Policy Directive Number 15, "Race and Ethnic Standards for Federal Statistics and Administrative Reporting." Adopted on May 12, 1977.
- Schwede, L. 1997. Report to the Office of Management and Budget on Third-person Reporting of Hispanic Origin and Race in a Group Quarter/Establishment Census. Available from the Statistical Research Division, Bureau of the Census.
- Schwede, L. and K. Ott. 1995. Children in Custody Questionnaire Redesign Project: Results from Phase 1. Available from the Center for Survey Methods Research, Statistical Research Division, U.S. Census Bureau.
- Schwede, L. and L. Moyer. 1996. Children in Custody Questionnaire Redesign Project: Results from Phase 2 Questionnaire Development and Testing. Available from the Center for Survey Methods Research, Statistical Research Division, U.S. Census Bureau.
- Schwede, L. and C. Gallagher. 1996. Children in Custody Questionnaire Redesign Project: Results from Phase 3 Cognitive Testing of the Roster Questionnaire. Available from the Center for Survey Methods Research, Statistical Research Division, U.S. Census Bureau.
- U.S. Department of Education. National Center for Education Statistics. *Racial and Ethnic Classifications used by Public Schools*. NCES 96-096 by Nancy Carey and Elizabeth Farris. Judi Carpenter, project officer. Washington, D.C. 1996.