

ASSESSING THE SIMPLIFIED ENUMERATOR QUESTIONNAIRE FOR CENSUS 2000

Diane F. Barrett and Michael Tenebaum
Diane F. Barrett, Bureau of the Census, Washington, D.C. 20233

Key words: *Questionnaire design; personal interviewing; nonresponse follow up*

I. INTRODUCTION

A team at the Census Bureau has been working to develop an enumerator questionnaire that will effectively collect demographic and housing data that is asked of nonrespondents in the Census 2000. The intent is to design a face-to-face instrument which will increase the quality of the data collected and reduce errors associated with personal interviewing. Specifically, the team was presented with the following mission:

To develop a simplified and effective data collection instrument (both enumerator and respondent friendly) for use in the United States Census 2000 Dress Rehearsal.

The team developed a prototype questionnaire, called the Simplified Enumerator Questionnaire or SEQ, to test in the 1996 Community Census. This paper presents an overview of the work that was done to develop the SEQ. It also summarizes item noncompletion rates as well as focus group and debriefing assessments of the SEQ.

II. BACKGROUND

In the 1980 Census, enumerators used the regular mail questionnaire to conduct nonresponse follow up and other field interviews. They were expected to reword the questions on their own to make them applicable for personal interviewing. There was little control over how enumerators asked the questions.

Starting with the 1990 Census, the Census Bureau first introduced an Enumerator Friendly Questionnaire for use during follow up operations. Although the wording was modified slightly for personal interviewing, the questionnaire was not that much different from the 1990 census questionnaire mailed to households. The Bureau found that fictitious persons and other types of coverage errors, such as duplication and omission of census residents, occurred at a higher rate on enumerator-filled forms than on mail returns [6].

During the planning for the 1995 Test Census, a work group was established to look at ways to improve the 1990 enumerator questionnaire. However, due to budget constraints and other resource restrictions, the 1995 enumerator questionnaire was once again modeled after the booklet style 1995 mail questionnaire. Observers of field

operations and a review of a sample of tape-recorded interviews during the 1995 Test Census found that enumerators reworded or omitted some of the questions, verified information rather than ask questions as worded, and made other changes to procedures which may have had implications for data quality. Although there were many reasons for what the enumerators did, part of it may have been attributable to the wording and the design of the enumerator questionnaire [7].

III. DEVELOPMENT OF THE SIMPLIFIED ENUMERATOR QUESTIONNAIRE

Based on the 1990 and 1995 census findings, the need to improve the design and use of the enumerator questionnaire for Census 2000 became a priority for the 1996 Community Census. The SEQ team took a zero-based approach to its work and was not constrained by the question wording, structure or format of the mailout questionnaire.

Wording and Format Development

The team made some fundamental changes, such as no longer requiring enumerators to ask sex as a separate question, asking for age only if the birth date was not known, using a household screener question for Hispanic origin and requiring use of a flashcard to simplify the collection of race data. The team tried to make the questions easy for the enumerator to read aloud, and simple for the respondent to understand.

Once the wording was determined, the layout and overall appearance of the form was developed. The team moved from the booklet approach used in 1995 to a single-sheet folded to four pages for the SEQ. This change allowed both sides of the SEQ to be data captured at the same time and also made the form appear much shorter to the respondents. Another significant change which saved time and eliminated transcription errors for enumerators was that the roster of names was entered once on the SEQ. On previous forms the roster was collected separately then names had to be transcribed to the appropriate person section of the questionnaire.

The team used a topic-based approach whereby each question is asked of all persons before moving onto the next question. Previous census questionnaires used a person-based approach whereby all questions are asked for each person before moving onto the next person.

Refer to Table 1 which compares some of the specific design features among the 1990 Census Enumerator Friendly Questionnaire, the 1995 Test Census Enumerator Questionnaire and the SEQ.

Iterative Testing

After the questionnaire wording and format were decided, the team moved into the iterative research and testing stage of its work. This stage involved several steps.

Cognitive Interviews - Several members from the team conducted cognitive interviews to gain some insight about how various respondents interpret the question wording [5].

Philadelphia Hot House - A small scale field test in the Philadelphia area was conducted in April 1996. Half of the interviews were completed with the 1995 enumerator questionnaire while the other half were conducted using the SEQ. According to the telephone debriefing session with the interviewers, they said that the SEQ took less time to complete and was easier to administer in the field than the 1995 version [4].

The item noncompletion rates for Hispanic origin indicated that marking a global box that represents all persons in the household instead of marking a box for each person on the 1995 enumerator questionnaire reduced item nonresponse. The item noncompletion rate was 2.1 percent on the SEQ and about 7 percentage points higher on the 1995 enumerator questionnaire.

1996 Community Census - Since the team had made tremendous progress in developing a viable instrument, it was decided to use the prototype SEQ to enumerate the nonresponse follow up cases in the 1996 Community Census.

IV. EVALUATION METHODOLOGY

The final stage of the team's work was to evaluate and analyze the SEQ used in the 1996 Community Census. The team planned several different evaluations using qualitative and quantitative methods to measure the effectiveness of the SEQ. This paper discusses the results from two qualitative methods to assess the effectiveness of the SEQ: focus group discussions and debriefing questionnaires [8]. The intent was to obtain detailed information about how well the SEQ worked in the field and learn about problems with the wording and design of the questionnaire. We also wanted to determine whether a newly designed form had an effect on data quality by examining item noncompletion rates and other data tabulations. The 1996 Community Census was conducted in selected tracts of Chicago and two American Indian reservations - Fort Hall, Idaho and the Pueblo of Acoma, New Mexico.

V. RESULTS

Item Noncompletion Rates

Table 2 provides the item noncompletion rates for the population questions as well as the coverage and tenure

items of the SEQ for each site. Table 2 also shows item noncompletion rates from the 1990 Census short form enumerator questionnaire for the specific Chicago tracts selected for the 1996 Community Census. The item noncompletion rates from the short form 1990 Census enumerator questionnaire for the same items except coverage are also provided at the national level. Since the rates are not estimates and were not derived from a sample, no statistical significance tests or associated standard errors were necessary for comparing item noncompletion rates between the 1996 SEQ and the 1990 Census enumerator questionnaire.

Enumerators in both Indian reservation test sites, especially Fort Hall, had little difficulty completing the population, coverage and tenure items on the SEQ. The noncompletion rates were consistently below the 1990 national average.

The SEQ item noncompletion rates for relationship, date of birth/age and Hispanic origin were substantially lower than the rates for these same items on the 1990 form for the same tracts in Chicago. Moreover, these rates were lower than the 1990 national average.

Enumerators had problems completing the coverage items more often than any other item on the SEQ. The rates were about the same, 10 percent. The SEQ contained two coverage questions-one to add persons to the household roster who are sometimes missed in the census, such as house mates and young children, and the other question was intended to identify persons erroneously reported as household members, such as persons living away in the military, at college or other special places on census day. Enumerators reported that they had to clarify, repeat and/or explain the purpose of the coverage questions at least some of the time. Respondents not understanding the questions may have contributed to the high item noncompletion rates for both coverage questions in the Chicago tracts. Since the need for coverage questions on the questionnaire wasn't introduced until the 1990 Census, no comparisons to the 1990 questionnaire can be made.

Detailed Results - Population and Tenure

Sex - The space for recording sex of each household member was incorporated into the roster column rather than a separate column. That is, instead of asking the sex item as a separate question, the enumerator was to mark the item by observation after obtaining each person's name.

The debriefing questionnaires and focus groups reported that the sex item was usually marked during the interview. However, crewleaders in Chicago reported that sex was sometimes missing on the questionnaires they edited, but they usually completed it based on name. Sex was missing for about four percent of the persons enumerated on the SEQ in Chicago which was about the same rate for persons enumerated on the 1990 enumerator form. The noncompletion rates at both Indian reservation sites were

less than two percent.

Relationship - During the development of the SEQ, the team discussed the need for a relationship flash card or some other way to simplify the selection of the proper relationship for each household member.

According to the debriefing questionnaires, most of the enumerators had to read the relationship categories at least some of the time before completing this item. Enumerators in all three sites said obtaining relationship for relatives were not as problematic as for nonrelatives. Enumerators usually had to read and/or probe for the nonrelative categories. The item nonresponse rate for the Chicago tracts was 2.7 percent which was the lowest item nonresponse rate on the 1996 SEQ and about three percentage points lower than the 1990 national average. Changing the wording from "*How is ... related to Person 1?*" to "*What is ... relationship to Person 1?*" may have increased the completion of this item.

Date of Birth and Age - Enumerators were to ask for age only for persons whose date of birth was not known. Enumerators reported that they had to ask for the age at least some of the time. The Chicago focus group participants reported that there was some resistance from respondents providing date of birth. They viewed the question as "too personal" and said it asked specific information that could be used by Immigration and other authorities. In the American Indian sites, enumerators expressed some concern about the sensitivity of asking elders their age since this may convey disrespect.

In spite of the sensitivity and personal nature of asking date of birth or age, the complete date of birth and/or the age was obtained for the majority of the persons enumerated in all three sites. Item noncompletion rates for the Chicago sites and the two Indian reservations were 5.4 percent, 0.5 percent and 1.7 percent respectively. It appears requesting the age only if the date of birth is not known reduces item noncompletion. On previous enumerator questionnaires, both the age and date of birth were required. The item noncompletion rate for the Chicago tracts was substantially higher on the 1990 enumerator questionnaire although comparable to the 1990 national average (8.8 percent) than the item noncompletion rate on the SEQ.

Hispanic Origin - One of the fundamental innovations of the SEQ was the addition of a household screener question to determine if anyone was of Spanish, Hispanic or Latino origin. If no such persons were in the household, enumerators marked the appropriate box and moved onto the race question. Otherwise, they asked the origin question to determine which Hispanic origin best described each person. Respondents were then to pick an origin category from the flash card.

The item noncompletion rate for the SEQ was substantially lower than the rate for the 1990 enumerator-filled forms in the Chicago tracts as well as the 1990

national average.

The majority of the enumerators reported that most respondents did not have any problems answering this question as a result of the origin screener item. Enumerators from the focus groups reported that respondents understood the various Hispanic origin categories and rarely were unable to choose one from the flash card. In the situations where there was some respondent uncertainty, it was usually because they did not understand what the question was asking, or they confused origin with race.

Table 3 provides the proportion of persons by response to the origin screener question. Table 3 also shows the number and percent of write-in responses.

In Chicago, about 47 percent of the persons enumerated during nonresponse follow up had one or more persons in their household who were Spanish, Hispanic or Latino. The proportion was much less for both the American Indian sites.

Of the enumerators with Spanish, Hispanic, or Latino households in their assignments, Chicago enumerators reported that respondents did not understand the various Hispanic categories and had difficulty in selecting the appropriate one.

About six percent of the persons enumerated in the Chicago tracts provided a write-in response. The majority of the write-ins were for some other Spanish, Hispanic or Latino origin.

Race - Historically, the race item has been problematic on census questionnaires. To give the respondent a point of reference for defining race as well as reducing the number of write-ins, the SEQ team decided to use a flash card to facilitate obtaining race information for each household member. The enumerator showed the flash card which listed the various race categories and asked which category best described each person's race.

As seen in Table 4, a specific race category was marked for more than 91 percent of the persons enumerated in each site. For those persons who were American Indian, the name of the enrolled or principal tribe was to be obtained. Respondents who chose "Other Asian or Pacific Islander" or "Some other race" category were asked the name of the specific race.

As expected, the majority of the persons enumerated in Fort Hall and the Pueblo of Acoma were classified as American Indian. About 98 percent of the write-ins in each of the American Indian sites were the name of their enrolled or principal tribe.

Chicago focus groups reported that Hispanic persons had difficulty distinguishing between origin and race, and it appears that respondents gave their Hispanic origin as their race, thus having the same information in both items. About 37 percent of the persons enumerated in Chicago had a write-in present and about 76 percent of the write-ins were for "Some other race." Because of the confusion

between origin and race, Hispanic persons may have chosen the "some other race" category from the flash card and gave the enumerator their Hispanic origin as their race. About 16 percent of the persons enumerated in Chicago with a specific race category marked also provided a write-in response.

Tenure - The housing question and response categories for determining tenure were slightly modified on the SEQ. Instead of asking about "this apartment or house," enumerators asked about this "place" and fewer words were used to describe the various tenure categories. The team was interested in learning whether the modifications had caused any problems in obtaining housing information.

Although most enumerators had no problems obtaining information, some enumerators reported that some respondents refused to answer because it is "none of our business." Other respondents were reluctant to answer the housing question because the enumerators told them that the interview only asked for information about the people living there, and not about whether they owned or rented their home. Enumerators felt that this question generated some hostility among their respondents. The 1996 tenure noncompletion rate was 7.4 percent which was about three percentage points higher than the 1990 national rate and the 1990 rate in the Chicago tracts.

V. CONCLUSION

Based on the qualitative data from the debriefings and focus groups, field staff as well as independent Census Bureau observers reported that the prototype SEQ used in the 1996 Community Census was an effective instrument that was easy to administer in the field. The new design seemed to work well and the overall format and flow, which involved a topic-based approach, was generally well received by respondents and was comfortable for enumerators. No major operational problems were identified.

The quantitative data indicate that a simpler designed questionnaire may have influenced enumerators' ability in completing several items. The high noncompletion rates for race, coverage and tenure, especially in the Chicago tracts, as well as the high item noncompletion rate for Hispanic origin item in the Pueblo of Acoma, are of concern. The inherent confusion between the origin and race items observed during this test probably was the largest contributing factor for the higher item noncompletion rates rather than the design of these questions. Note that the Hispanic origin item noncompletion rate for the 1990 Census enumerator-filled short forms in the same Chicago tracts were substantially higher than the SEQ. The Office of Management and Budget (OMB) is currently reviewing its statistical policy on race and ethnic standards for federal statistics and administrative reporting. The Census Bureau will comply with the results of OMB's review when

developing the specific questions for race and ethnicity for Census 2000.

The data also indicate that some items on the SEQ did exceptionally well such as the relationship, the date of birth/age and the Hispanic origin item. Not requiring both data and age and marking a global box that represents all persons in the household appears to be an improvement over the way these items were asked on previous questionnaires. Other items on the 1996 SEQ such as race, coverage and tenure need some modification, if we hope to minimize item nonresponse. Work has already begun on modifying the SEQ for retesting in the 1998 Dress Rehearsal for the 2000 Census.

REFERENCES

- [1] Barrett, Diane F., "1996 Community Census-Observation of the Nonresponse Follow-up Enumerator Training and Field Work in Chicago, Illinois." April 9, 1997.
- [2] Barrett, Diane F., "Philadelphia Hot House Test-Data Comparisons Between the 1995 Enumerator Friendly Questionnaire and the 1996 Simplified Enumerator Questionnaire," July 1997.
- [3] Borsa, Fred., "SEQ Team's Action Plan," internal Census Bureau memorandum, September 19, 1995.
- [4] Borsa, Fred., "Summary Report on the Debriefing of the Philadelphia Hot House Test," May 28, 1996.
- [5] Ciochetto, Susan., "Report on the Cognitive Research for the Simplified Enumerator Questionnaire." November, 18, 1996.
- [6] Griffin, Deborah, and Moriarity, Christopher., "Characteristics of Census Errors," Proceedings of the Section on Survey Research Methods American Statistical Association, 1992, Pages 512-517.
- [7] Griffin, Debbie and Cummings, Janet., "Improving the Enumerator Friendly Questionnaire for the 2000 Census," May 11, 1995.
- [8] Tenebaum, Michael, 1996 Community Census Results Memorandum No. 1, "Evaluation of the Simplified Enumerator Questionnaire Based on Debriefings and Focus Group Results," June 2, 1997.

ACKNOWLEDGMENT

The development and implementation of the SEQ has occurred through much effort of many individuals of the Census Bureau. The authors would like to acknowledge the SEQ Team for their hard work, dedication and endurance during the two years of design and evaluation of the SEQ. Special thanks to our Team Leader, Fred Borsa who kept us on track and moving forward. The authors who are

members of the SEQ Team salute the rest of the team: Sandy Chambers, William Chapin, Susan Ciochetto, Idabelle Hovland, Leitha Jamison-White, Dorothy Jenkins, Sherry Pollock, Glenn Schneider, Mark Wolfram, and Jeanne Woodward.

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

TABLES

Table 1. Design Features of Enumerators Questionnaires

1990 Census	1995 Test Census	1996 SEQ
6-pages, one sheet folded twice	8-page booklet	4-pages, one sheet folded in half
Information collected in a grid format. Person questions located in first column.	Information collected in person boxes. Person questions located in each person box.	Information collected in a grid format. Person questions located across the top row.
Person-based approach (ask all questions for a person before going to the next person)	Person-based approach (ask all questions for a person before going on to the next person.)	Topic-based approach (ask a question for all persons before moving to next question).
Abbreviated Introduction	140+ word Introduction	Abbreviated Introduction separated into three sections.
Roster listed separately, then names transcribed to person columns.	Roster listed separately, then names transcribed to person boxes.	Roster of names listed once, no transcription required.
Names collected last name first.	Names collected last name first.	Names collected first name first.
Separate question for Sex	Separate question for Sex with option to mark by observation.	Sex is marked by observation concurrent with listing of names.
Date of Birth and Age asked for each person.	Date of Birth and Age asked for each person.	Age asked only when Date of Birth is not known or incomplete..
Hispanic origin asked after the Race question for each person. No flash card.	Hispanic origin asked before Race for each person. Flash card available, use not required.	Household screener question for origin asked before Race. Required use of flash card for Hispanic persons.
Race question asked for each person. No flash card.	Race question asked for each person. Flash card available, use not required.	Race question asked only once for everyone in the household. Required use of flash card.
A second questionnaire was used for more than 7 persons.	A second booklet was used for more than 7 persons.	One page "continuation sheet" for more than 5 persons.
Notes section available. Callback record in address registers.	No notes section. Callback record in address registers.	Separate section to record callbacks on the form. Notes section available.

Table 2. Item Noncompletion Rates of the SEQ

Item	Chicago Tracts		Fort Hall	Pueblo of Acoma	1990 Census National
	1996 SEQ (%)	1990 EFQ (%)	1996 SEQ (%)	1996 SEQ (%)	1990 EFQ (%)
Sex	4.1	3.9	0.9	1.5	3.5
Relationship	2.7	9.3	0.3	2.0	6.1
Date of Birth/Age	5.4	8.3	0.5	1.7	8.8
Hispanic Origin	5.1	12.1	1.0	7.2	6.7
Race	8.9	7.9	0.4	2.5	4.7
Coverage-Add	9.6	NA	0.0	2.1	NA
Coverage-Delete	10.0	NA	0.1	2.4	NA
Tenure	7.4	4.2	1.1	0.7	4.0

Table 3. Hispanic Origin Screener

<i>P4. Is anyone we have listed Spanish, Hispanic or Latino? [] Other Spanish, Hispanic or Latino- Specify group</i>	Chicago Tracts		Fort Hall		Pueblo of Acoma	
	Num	Pct	Num	Pct	Num	Pct
Item nonresponse	445	11.9	2	0.2	11	3.8
No one in this household	1,556	41.6	703	85.8	271	93.8
Yes	1,741	46.5	114	13.9	7	2.4
Total write-ins	637	5.5	35	1.2	30	2.1

Table 4. Race

<i>P5. Which of these categories best describes each person's race?</i>	Chicago Tracts		Fort Hall		Pueblo of Acoma	
	Num	Pct	Num	Pct	Num	Pct
Item nonresponse	1,036	8.9	11	0.4	37	2.5
Category marked	10,657	91.1	3,039	99.6	1,416	94.3
“American Indian”	62	0.5	2,195	72.0	1,403	96.4
“Other Asian ”	63	0.5	0	0.0	0	0.0
“Some other race”	3,578	30.6	59	1.9	4	0.3
Specific race category	6,954	59.5	785	25.7	11	0.8
Write-in present	4,359	37.3	2,249	73.7	1,399	96.2