AN EVALUATION OF THE 1995 TEST CENSUS INTEGRATED COVERAGE MEASUREMENT (ICM) INTERVIEW: RESULTS FROM BEHAVIOR CODING

Nancy Bates
Cynthia Good

U.S. Bureau of the Census, Room 3127, FOB 4, Washington, D.C. 20233-9150
Project Hope, 7500 Old Georgetown Rd, Bethesda, MD.

Annual Meetings of the American Statistical Association,
Chicago, IL, August 3-8, 1996

of the Census Bureau.

KEY WORDS: CAPI, questionnaire design, reinterview

1. Introduction

The Integrated Coverage Measurement (ICM) process estimated the population of persons in the 1995 Test Census. The person interview component of the ICM was administered through a post-census CAPI interview. This instrument was a complex questionnaire constructed to serve multiple purposes. It was designed to gather an independent roster, conduct a thorough coverage interview, perform computer matching, resolve ICM/census roster discrepancies, and determine residency status. These competing goals, along with the complexities of a computer-assisted interview, presented numerous challenges for the questionnaire designers. The behavior coding explored in this paper provided valuable data from which to identify problems and begin the process of redesigning a simpler, more streamlined instrument.

This paper documents findings from a systematic review of a non-random sample of tape recorded ICM interviews conducted by a few interviewers at the Oakland and Louisiana test sites. These tape recordings served as the data for a questionnaire evaluation technique known as behavior coding, a questionnaire test methodology which systematically codes interviewer/respondent interactions. Despite some implementation problems, this method was useful in diagnosing problem question wordings, question order and the overall "flow" of the interview. Results from the behavior coding revealed many problem areas within the 1995 ICM interview. Practically every question analyzed surpassed the "problem" cutoff level. These data indicate that interviewers frequently modified question wordings or failed to read questions entirely. These findings were helpful in revising the ICM roster questions, re-ordering sections, and revamping the approach for resolving ICM/Census roster discrepancies and developing questions to determine residency status.

2. Methodology

2.1 Study Design

Field interviewers from the Oakland and Louisiana census test sites tape recorded a sample of ICM interviews which served as the basis for the behavior coding. Our research plan specified the involvement of 10 interviewers from Oakland and 10 from Louisiana. We instructed ICM local supervisors to select these interviewers from the graduating class of the first ICM training module. We did not intend this procedure to follow a formally randomized selection process.

The 20 selected interviewers were to begin tape recording as soon as they felt comfortable

with the laptops (after having completed about 1 week of interviews). Once they began taping, interviewers were to keep tape recording until they completed 15 taped interviews. Given these guidelines, we expected to obtain approximately 150 tape recordings from each site, for a total of 300 recorded cases.

By the closeout of ICM interviewing, the Center for Survey Methods Research (CSMR) had received 156 taped interviews from Louisiana but only 74 from Oakland. Eliminating unusable tapes[1], we were left with 122 Louisiana tapes and 64 Oakland tapes, far short of our requested 300 equally split between the two sites.

The recordings from Louisiana were produced by 14 different interviewers, most of whom recorded between 10 and 15 interviews each[2]. The Oakland tapes represented the work of only 8 different interviewers, 2 of whom produced over half of the tapes. The Oakland numbers obviously do not reflect the even case-load distribution that we had hoped for.

2.2 Limitations

Since the assignment of interviewers and interviews was not random, the generalizability of the behavior coding results is compromised. This is particularly true in Oakland, where the majority of data come from just 2 interviewers. We suspect that the interviewers who obtained so many tape recordings were above-average in terms of completion rates, refusal conversions, and overall performance. Consequently, they should not be viewed as representing the "average" interviewer's behavior in Oakland.

Another problem with the uneven case-load distribution is that the results become "weighted", in a sense, by those interviewers who provide more cases than others. For example, if one or two interviewers were responsible for a disproportionate number of recordings, then the behaviors of these interviewers can skew the overall distributions. If these interviewers tended to systematically exhibit a certain behavior (say, omit an introductory statement), then this will be reflected in the overall distributions when in fact, the behavior may characterize only a few interviewers. Alternatively, if interviewers who provided the most tape recordings were exceptionally "good" and had higher-than-average rates of exact/slight question readings, then this may mask the extent of other problems behaviors exhibited by interviewers who provided fewer cases.

Another limitation is the tape recording itself, which introduces unknown levels of bias into the research process. We suspect that interviewers may be more likely to follow the CAPI script, use flashcards, calendars, etc., when they know they are being tape recorded. Tape recording may also affect respondent behavior in unknown ways. Despite the limitations, we still felt that behavior coding was a worthwhile method for revising the ICM instrument. We believed that any problems uncovered for these "good" interviewers would be common to all interviewers.

2.3 Behavior Coding

Behavior coding is the systematic coding of the interactions between an interviewer and a respondent. Behavior coding is commonly used to assess whether interviewers have problems administering questions and whether respondents have difficulty comprehending questions, vocabulary, terms and concepts (Oksenberg, Cannell, and Kalton 1991; Morton-Williams and Sykes 1984; Marquis, Cannell and Robison 1971). This method is useful at indicating interviewer and/or respondent behaviors that may reflect problem questions, potential biases or inaccuracies in the data collection process. It's also a fairly inexpensive (although labor intensive) and relatively unobtrusive method compared to other pretest activities such as cognitive interviewing.

Two CSMR project staff members trained four experienced behavior coders to perform the coding according to project-specific procedures. Training and coding were conducted at the Hagerstown telephone interviewing facility. To behavior code the cases, coders listened to the tape recorded interview while simultaneously viewing the computerized trace file for the case. Trace files allowed the coders to "play back" the CAPI interview step-by-step, exactly as it occurred in the field.

This paper focuses on two components of the behavior coding scheme: question-asking codes and response codes. The first step, coding the initial question-asking behavior of the interviewer, is important because if many interviews show a deviation in wording on a particular question, it usually indicates that a question is poorly worded. The major categories for interviewer question asking behavior are as follows[3]:

Question Asking Codes:

Exact Wording or Slight Change - The interviewer asked the question exactly as written or with only slight modifications that did not change the meaning of the question.

Major Change in Question Wording - The interviewer administered the question with major changes to the scripted question wording that altered the intended meaning of the question (such as omitting key words, phrases, or dates or by paraphrasing).

Verification - The interviewer verified or repeated relevant information that the respondent had provided earlier, in place of asking a specific question.

Omission - The interviewer entirely omitted (answered without reading) an applicable question.

After coding the interviewer's presentation of a question, coders recorded the respondent's

initial response to it. Coding respondent behavior is important for determining whether respondents are having difficulty understanding the meaning of questions and for identifying sensitive questions. Response codes are as follows:

Response Codes:

Adequate Answer - The respondent provided an adequate answer that met the objective of the question.

Inadequate Answer - The respondent provided an answer that did not meet the objective of the question and required additional probes to ascertain an adequate answer.

Break-in - The respondent interrupted with an answer before the interviewer finished reading the question.

Clarification - The respondent asked the interviewer to clarify the meaning of a particular question or concept, or asked for a repeat of the question.

Other Respondent Behavior - The respondent did something not covered by one of the other response codes (assumed non-verbal response, garbled recording, tape drop-out, etc.).

Whenever a major modification or inadequate answer occurred, coders recorded a brief note to indicate the specific modification or content of the inadequate answer.

Research indicates that behavior coding can be used to evaluate questions, but in order to do so, the coding must be reliable--that is, each coder must apply the same codes to the same behaviors. As an evaluation of the coders' grasp of the materials presented in training and to measure inter-coder reliability, we computed the reliability statistic kappa based on the same case coded individually by all four coders. We conducted inter-coder reliability tests at two different times, once immediately following training before full-scale coding began, and a second time, using a different case, about two-thirds of the way through production.

Values of kappa above 0.75 are said to represent excellent agreement, and values from 0.40 to 0.75 represent fair to good agreement beyond chance (Oksenberg, Cannell & Kalton, 1991). We computed separate kappa statistics for question asking codes and for response codes. Within each of these categories, however, the kappa statistic represents the reliability among coders for all codes, not for individual codes within the category. We also calculated an overall percent agreement rate among the coders across all questions, interviews, and codes.

For each reliability test, we generated six kappa statistics in each category, one for each pair of coders[4]. For the first reliability test, the kappa for question asking codes ranged from 0.57 to 0.73, and for response codes ranged from 0.59 to 0.82. The overall percent agreement rate among the coders was 83 percent. The second reliability test yielded kappas for question asking codes ranging from 0.67 to 1.0 and response codes ranging from 0.38 to 0.89. The overall percent agreement rate among the coders was 87 percent. These results indicate that the reliability among our coders was within an acceptable range.

Following standard practice[5], we used 15% as a general guideline to indicate problem questions; that is, if 15% or more of the question readings had "problem" behaviors (e.g. a major change) then this indicated a significant level of the problem. We applied a more stringent cutoff in the case of question omissions (10% or higher was considered significant) because we felt this behavior was an obvious indicator of severe design problems (the question was perceived by interviewers as too sensitive, redundant, illogical, etc.).

3. Results

3.1 The Independent ICM Roster

The first section of the ICM interview attempted to obtain the most accurate and thorough household listing possible. The intent was to obtain the "true" Census Day roster. It begins by asking for all persons living permanently or staying temporarily at the sample household on Census Day and is followed immediately by a battery of six probes meant to stimulate respondents to consider people commonly left off household rosters. Tables 1A and 1B illustrate the question wordings and summarize the interviewer and respondent behaviors during the roster section.

Table 1A indicates that the roster section had both a high incidence of major wording modifications and relatively frequent question omissions. To begin with, interviewers made major changes to the calendar presentation procedures (CALENDAR) in 39% of the interviews. In most cases where a major change occurred, interviewers omitted the day and year and mentioned "March 4" only. Despite the paraphrasing, it was reassuring to see that most interviewers at least attempted to orient the respondent to the critical date. The flashcard calendar probably aided this behavior.

The combination of the roster question (NameA) immediately followed by the listing instruction (NameB) was apparently not a very smooth start to the interview. Interviewers made significant modifications to the roster question 20% of the time and over half the time for the listing instruction. Common changes to the roster question included omission of the reference date and the phrase "staying here." A portion of this is probably due to the high respondent break-in rate (27% Table 1B).

Interviewers frequently paraphrased the listing order instruction (29% Table 1A) by shortening it or routinely omitting the last sentence. In close to 20% of the interviews, the instruction was omitted altogether (Table 1A).

These interviewer behaviors may have contributed toward some of the inadequate respondent answers at NAMEB (see Table 1B) such as "it's just me and my little boy". Such answers do not meet the objectives of the question (a list of names). Other inadequate responses such as "I own my own home" resulted from the order instruction, which sidetracked respondents from their listing task.

In close to one-quarter of the cases, interviewers modified the introduction to the roster probes (INTRO) so that the meaning was changed (Table 1A). Most times it was paraphrased into something like "I have a few questions to make sure you didn't forget anybody," but some interviewers seemed to be warning the respondent of the probes to follow by making statements like "I know that this sounds redundant but they ask me to ask these questions ..." or "there's a few questions here, they may seem a bit redundant but they're designed for a reason." More unsettling is that in 11% of the interviews, the introduction was skipped altogether. We consider this a serious error since this statement serves as the only explanation of the critical roster review that follows.

Table 1A suggests that interviewers frequently modified the wordings of the first three roster probes (A1@a, A1@b, A1@c). All three questions surpassed the 15% cutoff for major modifications. Some of these wording changes were due to reading only a partial list of the examples - e.g., "have I missed anyone temporarily away or on a business trip?" Other common errors were the omission of reference dates and the omission of clauses at the end of a probe e.g., "in a general hospital", "live-in employee", or "child away at boarding school". Some interviewers tended to offer probes in a biased negative manner e.g., "no one off shore coming in for the weekend or anything like that?", "no roommates or foster children?" Perhaps the most common major modification was to collapse several probes into one by simply picking off one or two examples from each.

The frequency of omissions for the probes, ranged from 8% for the first, to double that amount for the last (16%). This may have been interviewers' response to respondent break-ins during earlier probes. It was apparent in some households (and in particular, one-person households) that the respondents perceived the probes as a redundant nuisance. This behavior helps illustrate a basic difficulty in the ICM interview. In order to uncover census omissions and determine erroneous inclusions, interviewers must apply intensive probing questions. These probes must be applied in each interview, but result in uncovering ICM/census discrepancies for only a small percentage. Consequently, interviewers are faced with a "needle-in-a-haystack" phenomena which requires patience and a good understanding of the survey's intent.
In performing the last rostering task, the review of names (Names@1), interviewers also frequently modified the item wording (30%), probably due in part again to respondent break-ins (16%). These interruptions discouraged interviewers from reading the last "have

I missed anyone" probe. In general, the high frequency of major modifications, omissions, and break-ins for this section make it clear that interviewers did not administer the ICM rostering section as its designers intended.

3.2 Roster Recommendations

To improve the listing instruction, we have moved the second part (NAMEB) to become a separate question that follows the initial roster task in the revised instrument. It has also been shortened: "in whose name is this house/apt. owned or rented?" Interviewers then simply flag the line number of the appropriate person. The calendar used to orient respondents to Census Day has been retained in the next instrument. Drawing the respondent's attention to the calendar is critical since the time lag between the census and ICM interview will be between 3-4 months.

It is obvious from the interviewer behaviors that the extensive list of coverage probes was not administered as written. Sensing that the list is redundant, interviewers routinely shortened the questions, combined several questions into one, or simply omitted probes altogether. Interviewers also sometimes failed to provide an adequate context for the probes by omitting the introduction in over 10 percent of the interviews. Since the quality of the independent roster is perhaps the most important component of the ICM interview, we recommended a completely new rostering technique for the next ICM instrument.

A new rostering alternative has been tested as part of the next ICM test cycle. The new approach guides respondents through the cognitive task of reconstructing their Census Day household roster using a process that differs from what they used in the census and also acknowledges the substantial time lag since Census Day. The approach first inquires about persons who stayed at the sample unit on the night before the ICM interview. For each of these individuals, the interviewer determines whether the person also stayed in the unit on Census Day. Next, the respondent is asked a series of cues to aid recall of additional persons staying at the unit on Census Day who were not staying there the previous night (for example, persons who have moved away or persons away temporarily).

This approach (known as the retrospective approach) was designed to re-create Census Day rosters by first using information most accessible in memory (who was there last night) and then methodically working backwards (Biemer 1995; Sudman, Bradburn and Schwarz 1996). The resulting set of roster probes are shorter and more content-varied than the follow-up questions found in the '95 instrument. These efforts are meant to reduce the cycle of impatient interviewer interruptions and restructure the roster section by avoiding repetitive probes asked by rote.

3.3 Behavior at the Match Screen

Following a section of demographic questions (sex, age, D.O.B. relationship, ethnicity and

race -- not evaluated here), the ICM interview next initiated a computer-matching routine. Using a matching algorithm based on sex and date of birth, the computer displayed the matches (and non-matches) between persons listed during the ICM interview and persons listed in the original census report. The path of the instrument from this point on was driven by the outcome of the matching procedure. If the match indicated "new" persons on the ICM roster who were not listed during the census, or "missing" persons listed in the census but not in the ICM, the instrument branched to questions which obtained living situation and residency information for the non-matched persons.

In both cases, interviewers assigned a reason code for the discrepancy by asking an open-ended question about each non-matched person. For new persons listed during the ICM interview but not in the census the following question appeared:

> "I have a few more questions to ask you about the people you mentioned today. Can you think of any reason why [NAME] was not mentioned on the census form?"

Similarly, for missing persons listed at the address during the census but not mentioned during the ICM interview, the instrument asked:

> "I have a few questions to ask you about the people listed on the census form. [NAME] was not mentioned today. Can you think of any reason [he/she] was listed on the census for this address?"

Based on the respondent's answer, interviewers selected from a list of codes, some of which reflect living situations that might explain the discrepancy. The codes were meant to help interviewers select answers that would classify a non-matched person's residency status according to census residence rules. For "new" ICM persons, interviewers selected one of 16 reconciliation codes; for un-matched census persons, interviewers picked from a list of 19 (see attachment A for code lists).

Table 2A indicates that for both questions, omissions were frequent. When extra persons were left over on the ICM roster (B4@1), the question was skipped 30% of the time; when persons were left over from the census roster (C5@1), the question was omitted over half the time. The infrequent occurrence of these questions resulted in a fairly small number of codes (especially respondent codes) making inferences somewhat unreliable. Nonetheless, we make several observations about these questions based upon general questionnaire design principles.

First, the list of response codes for both questions is very long: 16 possible codes for question B4@1 and 19 codes for C5@1. Additionally, neither list's options are mutually exclusive. Further, because the codes attempt to correspond with official census "residence rules" they do not match typical answers provided by respondents, such as "Yes, I filled it out" or "No, I can't." Consequently, respondent answers did not always fit

neatly into one of the precoded response categories.

Part of the problem is that the response codes reflect an open-ended format yet the questions themselves are posed as yes/no questions. This problem is further accentuated because respondents are not shown or read the code list and therefore don't have any context within which to frame their answers. As a result, interviewers had to use a variety of strategies to elicit an adequate answer. Some restructured the question so that it specifically asked or verified where the unmatched person was on Census Day. Many ignored the codes and went straight for the "Other" category so they could record the respondents' verbatim answers. In fact, 86% of the B4@1 answers were recorded under "Other" while 32% of C5@1 were "Other" (West 1995). Because the instrument does not follow a structured question-and-answer process at this point, interviewers were left to their own devices to obtain an answer, which often required a great deal of respondent/interviewer interaction.

3.4 Match Screen Recommendations

To improve the flow of the instrument, we have moved the match screen to the end of the interview. This keeps the match screen "down-time" from disrupting the course of the regular interview. It also seems more natural to conduct the ICM/census comparison after the ICM interview is essentially completed. Based somewhat upon the behavior coding results (but perhaps more so on common sense questionnaire design principles) we decided to eliminate this open-ended section entirely. The process of selecting from a long list of non-mutually exclusive answer codes did not work well toward resolving roster discrepancies. Interviewers frequently omitted the questions and tended not to use the response categories choosing "Other" instead. To avoid this, we have implemented a structured question-and-answer approach to gather information necessary for assigning residency and resolving the ICM/Census discrepancies. This series of questions have now been integrated into the residency questions (described in detail below.)

3.5 The Residency Questions

The last section of the '95 instrument established or confirmed the correct Census Day residency status for all persons enumerated in both the ICM and the census. For persons who match up across the ICM interview and census, the series began with a screener question:

> "Some people have more than one place where they stay. Did [NAME] live or stay anywhere else in the past year?"

Positive responses were followed up by the questions about the type of other place, amount of time spent there, where the person was on Census Day, and where the person's "main residence" was. In the '95 instrument these questions followed the match screen.

In cases where all persons matched up across the ICM and the census, these questions were administered on a person-by-person basis, beginning with the householder. In cases with unmatched people, the residency questions were first asked of the unmatched persons and then of everyone else. Except for minor wording differences, the questions were essentially identical across both situations. Tables 3A and 3B represent a combination of all the residency questions, regardless of the order in which they were administered.

A quick review of table 3A suggests a variety of problems with this battery of questions. All of the questions surpassed the 15% rule of thumb cutoff on one or more "problem" interviewer behavior codes.

To begin with, interviewers frequently modified the wording of the screener question (OTHER PLACE) or simply omitted it. Major changes resulted when interviewers omitted the introductory sentence or the phrase "in the past year." These behaviors reflect questionnaire design errors since the context of the question was often altered (by omitting the introduction) or answers were filled without asking. Since determining residency is one of the primary objectives of ICM, these design errors may have serious consequences.

Interviewers tended to omit the next question, type of place (TYPE PLACE), or verify answers. It is possible that omissions occurred when interviewers obtained the information in the immediately preceding question, which allowed him/her to answer without asking. Our coding cannot confirm this, but the code frequencies are consistent with omission and verification levels reported for the same item in the pretest where this proved to be the case (see Bates and Kindred-Town 1995). Both the open-ended nature of this question coupled with the difficulty of using flashcards may have contributed toward departure from the questionnaire script.

The third question in this series (TIME THERE) was also open-ended, although nine different time period "examples" were displayed below the question to act as guides for the interviewer. Approximately one-third of the time this question was modified such that the meaning was changed (a major modification); in close to another quarter of the cases it was omitted completely. The most common modifications were to omit the introductory sentence or to completely paraphrase the wording into things like "do you know what time he was away from this house here" or "What...nine months of the year?"

Some respondents had trouble answering this question as well. Thirty-seven percent provided inadequate answers such as "he stays there all the time, but he comes here" or "every now and then." It is difficult to tell from the behavior coding whether interviewers usually read the examples or not. (It is also unclear from looking at the instrument whether or not they were supposed to).

The coding also reflects high levels of problems with the last three items in this series (MAR.4, MAIN RES, 3RD PLACE). Here, the changes weren't so much wording

modifications as they were omissions (15-30%). Where omissions occurred, it's again possible that interviewers had obtained sufficient information during previous portions of the interview (perhaps at the match screen) to correctly answer without asking. If this was the case, then these data suggest the questions are somewhat redundant. An alternative explanation is that interviewers don't recognize the distinct importance of each question and therefore tend to data-fill later questions depending upon the answer(s) to earlier ones. One example would be to assign residency on the basis of only one question -- where the person was on Census Day. Knowing this, interviewers may be encouraged to self-fill other answers they believe are superfluous and can be logically inferred.

3.6 Residency Question Recommendations

The residency questions have been simplified and moved to a different section of the interview. The new questions have been placed just before the match screen. This makes it possible to determine residency status for all persons rostered during the independent ICM interview, allowing most interviews to end immediately after the match screen. The only cases requiring additional questions past the match screen are those cases with persons left unmatched from the census roster and whole-household non-matches. Cases with persons left over on the ICM roster do not have to continue past the match screen because everyone's Census Day residency information is already ascertained.

To reduce perceived redundancy and hopefully decrease interviewer omissions, the new residency questions have been broken into three distinct sets: 1) within-household in-movers after Census Day, 2) persons residing in group quarters housing on Census Day and 3) persons with multiple residencies (non-group quarters) on Census Day. This allows each question to be asked only once, in a group-manner, rather than on a repetitive and time-consuming person-by-person basis. We hope the content of these questions are now sufficiently different from one another so that interviewer omissions are discouraged.

Residency status for persons with multiple residencies are now based on a simple concept. According to the recently revised census residence rules, this is determined by where the person "lives and sleeps most of the time" (Rolark 1995). In cases where the respondent is not sure, a similar rule-based process is now followed with no write-ins allowed. This restructuring moves away from the old method which attempted to categorize discrepancies according to specific living situations (e.g., children in joint custody, live-in employees, commuter workers, etc.), and instead applies "umbrella" residence rules for persons with more than one residence. This new format replaces the open-ended reconciliation questions used in the '95 instrument.

The behavior coding presented in this paper served as a useful diagnostic
toward the larger goal of building the next generation ICM instrument. During the next ICM test cycle, the revised instrument will undergo further rounds of coding plus other pretest methodologies such as interviewer and respondent debriefings and cognitive interviews.

These activities should continue to improve the accuracy of data collected during the ICM interview, a critical component of the Census 2000.

References

Bates, N. and C. Good (1995)
"The Integrated Coverage Measurement Evaluation of Tape Recorded Interviews with Behavior Coding: Project 2," DMD 1995 Census Test Results Memorandum Series, No. 18, December 15.

Bates, N. and M. Kindred-Town (1995)
"The November Integrated Coverage Measurement (ICM) Test: Results from Behavior Coding of ICM Person Interviews," U.S. Bureau of the Census report, Center for Survey Methods Research, Statistical Research Division, April 10, 1995.

Biemer, P. (1995)
"Draft Comments on ICM Evaluation Report Frames Related to Reconciliation Bias," Research Triangle Institute memorandum to H. Woltman, September 7, 1995.

Ellis, Y. (1994)
"Categorical Data Analysis of Census Omissions," U.S. Bureau of the Census, DSSD 1990 REX Memorandum Series #PP-10, July 26, 1994.

Fowler, F. (1992)
"How Unclear Terms Affect Survey Data," Public Opinion Quarterly, Vol. 56, 218-231.

Marquis, K., C. Cannell and S. Robison (1971)
"Report on a Tape-Recording Analysis of Interviewer-Respondent Interaction in three Urban Areas. Chapters 10-12, in J.B. Lansing el al. (eds.) Working Papers on Survey Research in Poverty Areas, Ann Arbor: Institute for Social Research, 1971.

Morton-Williams, J. and W. Sykes (1984)
"The Use of Interaction Coding and Follow-up Interviews to Investigate Comprehension of Survey Questions," Journal of Market Research Society, 26, 109-127.

Oksenberg, L., C. Cannel, and G. Kalton (1991)
"New Strategies for Pretesting Survey Questions," Journal of Official Statistics, Vol. 7, No. 3. pp. 349-365.

Rolark, S. (1995)
"Residence Rules for the 2000 Decennial Census," U.S. Bureau of the Census memorandum for A.J. Norton from S. J. Rolark, October 20, 1995.

Sudman, S., N. Bradburn and N. Schwarz (1996)
Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. Jossey-Bass: San Francisco, 1996.

Sweet, E. (1994)
"Roster Research Results from the Living Situation Survey," U.S. Bureau of the Census, 1994 Annual Research Conference Proceedings, pp. 415-433, March, 1994.

Wellens, T. and E. Gerber (1996)
"ICM Cognitive Evaluation," 1996 Census Test Memorandum Series Chapter IP-QD-1, June 28, 1996.

West, K. (1996).
"ICM Evaluation Project 4: Response to Coverage Probes in the ICM Person Interview," DMD 1995 Census Test Results Memorandum No.21, January 29, 1996.

Table 1A.
Roster Questions - Interviewer Reading Behavior

(NOTE: Only the major codes are shown in the tables - row totals may not sum to 100%)

	N	Exact/Slight	Major Change	Omitted	Verified
Calendar	188	56%	39%	5%	0%
NameA	186	68	20	11	0
NameB	182	29	52	19	0
Intro	184	66	23	11	0
A1@a	185	58	32	8	2
A1@b	185	67	22	9	2
A1@c	185	62	25	10	3
A1@d	184	73	14	13	1
A2@e	184	73	10	15	1
A2@f	185	72	11	16	1
Q1name	5	80	20	0	0
Names@1	185	57	30	8	6

Calendar: Please look at this calendar. The critical date is Saturday, March 4, 1995.

NameA: What are the names of everyone who was living here permanently or staying here temporarily on March 4, 1995?

NameB: Please start with the name of the household member, or one of the household members, in whose name this house or apartment is rented, being bought or owned. If there is no such person, start with any adult household member.

Intro: We are trying to make sure that we count everyone in the census and count them at the right place. I am going to ask a few questions about people we sometimes miss.

A1@a: Have I missed, anyone who usually lives here, but was temporarily away, spending the weekend with a parent, on a business trip, on vacation, or in a general hospital on March 4?

A1@b: Any housemate, roommate, foster child, roomer, boarder, or live-in employee?

A1@c: Have I missed any young children, or babies born on or before March 4, or a child away at boarding school?

A1@d: Anyone staying here most of the week while working, even if that person has a residence somewhere else?

A2@e: Have I missed anyone who lived here on March 4, but has since moved out?

A2@f: Anyone who stayed here on March 4, who has no other place to stay?

Q1name: What is the person's name? [Asked if "yes" to A1@a-A2@f]
Names@1:I have listed [names]. Have I missed anyone?

Table 1B.
Roster Questions - Respondent Behavior

(NOTE: Only the major codes are shown in the tables - row totals may not sum to 100%)

	N[1]	Adeq.	Brk-in	Clar.	Inad.	Other
NameA[+]	15	47%	27%	13%	13%	0%
NameB	134	75	1	8	15	2
A1@a	166	78	11	1	1	10
A1@b	165	79	9	1	1	10
A1@c	162	78	10	1	1	9
A1@d	161	83	4	2	1	10
A2@e	156	87	1	1	1	9
A2@f	156	87	3	1	1	9
Q1name	5	100	0	0	0	0
Names@1	148	72	16	0	1	11

[+] These reflect cases where NAMEB was omitted and respondents answered at NAMEA.

[1] Because of interviewer omissions, the N's in table 1B and all subsequent respondent behavior tables are not the same as the N's in the corresponding interviewer behavior table; if a question reading was omitted, a respondent behavior was not coded.

Table 2A.
Reconciliation Questions - Interviewer Reading Behavior

(NOTE: Only the major codes are shown in the tables - row totals may not sum to 100%)

	N	Exact/Slight	Major Change	Omitted	Verif.
B4@1	56	30%	13%	30%	25%
C5@1	16	31	6	56	0

Table 2B.
Reconciliation Questions - Respondent Behavior

(NOTE: Only the major codes are shown in the tables - row totals may not sum to 100%)

	N	Adeq.	Brk-in	Clar.	Inadeq.	Other
B4@1	35	60%	0%	6%	3%	14%
C5@1	7	57	0	0	14	0

B4@1: "I have a few questions to ask you about the people you mentioned today. Can you think of any reason why [NAME] was not mentioned on the census form?

C5@1: "I have a few more questions to ask you about the people listed on the census form. [NAME] was not mentioned today. Can you think of any reason he/she was listed on the census for this address?

Table 3A.
Residency Questions - Interviewer Reading Behavior

(NOTE: Only the major codes are shown in the tables - row totals may not sum to 100%)

	N	Exact/Slight	Major Change	Omitted	Verif.
OTH. PLACE	475	53%	29%	12%	2%
TYPE PLACE	33	21	18	36	21
TIME THERE	45	33	36	24	7
MAR.4	46	59	2	20	20
MAIN RES.	47	70	0	15	15
3RD PLACE	47	62	6	30	2

Table 3B.
Residency Questions - Respondent Behavior

(NOTE: Only the major codes are shown in the tables - row totals may not sum to 100%)

	N	Adeq	Brk-in	Clar.	Inadeq	Other
OTH. PLACE	417	80%	6%	4%	3%	7%
TYPE PLACE	21	76	0	5	10	5
TIME THERE	35	43	0	9	37	3
MAR.4	38	74	0	3	3	8
MAIN RES.	40	78	3	5	5	3
3RD PLACE	33	88	3	6	0	0

Oth. Place: Some people have more than one place where they stay. Did [NAME] live or stay anywhere else in the past year? [If yes, continue]

Type Place: And what was that place? (show flashcard)

Time There: How much time did he/she spend there or when did he/she stay there?

Mar. 4: At which residence was [NAME] staying on March 4?

Main Res.: Which place do you consider to be [NAME'S] main residence?

3rd Place: Is there any other place where [NAME] stayed in the past year?

Attachment A

Reconciliation Codes for ICM non-matched persons

a) **College.** Lives away at college.
b) **Works Elsewhere.**
c) **Armed Forces.** Away in the Armed Forces.
d) **Rehabilitation or Substance Abuse Center.**
e) **Institution.** Such as prison, mental hospital, nursing home, or chronic care facility.
f) **New Baby after March 4.**
g) **Moved in after March 4.** Did not live here on March 4.
h) **Temporarily Away.** Such as vacation, visiting family, business trip, or in a general hospital.
i) **Boarding School**
j) **Here on March 4 while looking for a place to live.**
k) **Born on/before March 4.**
l) **Foster/young child living here on March 4.**
m) **Roomer/boarder living here on March 4.**
n) **Live-in Employee.**
o) **Moved Out after March 4.**
p) **Other Reason.**

Reconciliation Codes for Census non-matched persons

a) **College.** Lives away at college.
b) **Works Elsewhere.**
c) **Armed Forces.** Away in the Armed Forces.
d) **Rehabilitation or Substance Abuse Center.**
e) **Institution.** Such as prison, mental hospital, nursing home, or chronic care facility.
f) **New Baby after March 4.**
g) **Moved in after March 4.** Did not live here on March 4.
h) **Forgot to Mention Person**
i) **Temporarily Away.** Such as vacation, visiting family, business trip, or in a general hospital.
j) **Boarding School**
k) **Here on March 4 .**
l) **Born on/before March 4.**
m) **Foster/young child**
n) **Roomer/boarder**
o) **Live-in Employee.**
p) **Death.**
q) **Moved Out after March 4.**
r) **Lived in another housing unit in this block.**
s) **Other Reason.**

1. Unusable tapes included noninterview cases (vacants, refusals, etc.), tapes without identifiers, blank tapes, tapes with garbled audio and interviews in foreign languages.

2. Three of the Louisiana interviewers only taped between 2 and 4 interviews, presumably because they quit or were let go before their 15 tape quota was achieved.

3. The full list of respondent code categories also included multiple verifications and question collapsing; interviewer codes also included qualified answers and a "don't know" category. Since only the basic codes are presented here, row percentages may not always sum to 100.

4. Kappa statistics were calculated based on the full set of respondent and interviewer codes. Consequently, the estimated extent of agreement for the major categories of behavior is probably somewhat conservative.

5. The 15% cutoff is a standard index applied in other behavior coding research studies (see Oksenberg, Cannell, Kalton 1991; Fowler 1992).