RESEARCH REPORT SERIES
(*Survey Methodology #2002-04)*

ACS/CATI Person-Based/Topic-Based Field Experiment

Jeffrey C. Moore and Laureen Moyer

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: February 19, 2002

ACS/CATI Person-Based/Topic-Based Field Experiment[1]

Jeff Moore and Laurie Moyer
Statistical Research Division
U.S. Census Bureau

Abstract: The CATI interviews conducted among ACS-National sample mail nonresponse cases in October and November of 1997 included an experimental test of two different instrument structures — the traditional "person-based" approach versus a new "topic-based" design. A person-based interview in essence completes an entire interview for the first eligible household member, then returns to the beginning and completes an interview for the second person, and so on through all eligible persons. In contrast, a topic-based design gathers data on one "topic" for every person and then proceeds to the next topic, in effect making only one "pass" through the instrument. This paper presents the results of that experiment, comparing the performance of the two instrument designs on multiple dimensions: response/refusal rates, length of interview, assessments of interviewers and their supervisors, respondent evaluations, behavior coding of interviewer/respondent interactions, item nonresponse, and data outcomes including response distributions and within-household response consistency. In most respects, we find the topic-based design clearly superior to the person-based design.

Keywords: questionnaire design, household surveys, nonresponse, data quality, interviewers' evaluations, respondents' evaluations

1. Executive Summary

The CATI interviews conducted among ACS-National sample mail nonresponse cases in October and November of 1997 included an experimental test of two different instrument structures — the traditional "person-based" approach versus a new "topic-based" design. A person-based interview in essence completes an entire interview for the first eligible household member, then returns to the beginning and completes an interview for the second person, and so on through all eligible persons. In contrast, a topic-based design gathers data on one "topic" for every person and then proceeds to the next topic, in effect making only one "pass" through the instrument. Interview cases were assigned to one or the other instrument treatment at random. In addition, interviewers were assigned to one of two groups, one of which conducted all person-based interviews in October before switching to topic-based in November; the order of instrument treatments was reversed for the two test months for the other interviewer group.

---

[1]This paper was originally prepared as the final project report for the ACS/CATI questionnaire design experiment, dated July 29, 1998. Note that in this version of the paper, all tables have been moved to the back. Contact: Room 3133-4; Center for Survey Methods Research; Washington, DC, 20233-9150; phone: (301) 457-4975; fax: (301) 457-4931; e-mail: jeffrey.c.moore@census.gov.

A working group comprised principally of staff from the Continuous Measurement Office (CMO) and CSMR/SRD developed a set of evaluation procedures for a comparison of the two instrument designs, the essential results of which are as follows:

Interview Response/Refusal Rates The response rate for the topic-based treatment, 60.5%, was 4 percentage points higher than the rate for the person-based treatment. This significant advantage to the topic-based treatment appears to have resulted primarily from a reduction in refusals, which occurred almost 3 percentage points less frequently than in the person-based treatment (13.0% versus 15.9%, respectively). Virtually all refusals occur during pre-interview "negotiations," well before the structure of the interview is even potentially apparent to respondents. Thus, we conclude that the refusal rate difference between the instrument treatments in this test must arise from differences in interviewers' behavior in the face of similar base rates of respondent reluctance to participate. (See section 3.1.1, below, for details.)

Interview Length Overall, and excluding one-person households (in which the two instrument structures were indistinguishable), topic-based interviews were significantly shorter than person-based interviews on average. The difference in length — about 2 minutes — was concentrated in and quite consistent across two-, three-, four-, and five-person households; households consisting of six or more persons showed no significant interview length differences by instrument type. (3.1.2)

Supervisor Debriefing Questionnaire The 6 ACS-CATI supervisors provided feedback on various aspects of the person-based and topic-based instrument designs via a debriefing questionnaire. Overall, the supervisors were about evenly split in their preference for one design or the other (although they did admit to a possible bias towards the person-based instrument because of their experience and familiarity with it). Supervisors tended to view some specific aspects of the topic-based design negatively ─ for example, the use of first names ("too friendly ─ less professional"); the way that D's and R's had to be entered in situations involving an unknowledgeable but otherwise eligible proxy respondent (supervisors expressly requested the capability, in either instrument, to skip over, and call back later, a person whose data could not be provided by a proxy); and the greater difficulty in concentrating on one person's data. On other dimensions they noted the advantages of the topic-based design ─ shortened interview time; "much happier respondents;" and a less repetitious interview, especially for large households. (3.2.1)

Interviewer Debriefing Questionnaires Interviewers' responses to debriefing questionnaires administered near the end of each interview month indicated a consistent and strong preference for the topic-based instrument. Interviewers rated the topic-based instrument more favorably than the person-based instrument after the first interview month, before they had experience with both instruments, and their preference generally became even more marked after the second month, when they could directly compare the two. They were especially favorable to the topic-based instrument in larger households and for more reluctant respondents. In general, interviewers felt that the topic-based instrument made it easier to establish rapport with respondents. According to the questionnaire responses, the person-based instrument was not the interviewers' preferred instrument even in households consisting of non-relatives, the primary circumstance in which it was expected to out-perform the topic-based instrument. (3.2.2)

Interviewer Debriefing Focus Group In a focus group held near the end of the second month's interviewing, interviewers voiced a very strong and consistent preference for the topic-based instrument. Interviewers felt they spent less time and less energy conducting the topic-based interview, and that the interview was more conversational than the conventional person-based design. They thought that respondents had more confidence in them because the instrument allowed them to ask the questions in a less stilted and inflexible manner, which made them appear more engaged in and in control of the interview interaction. Several interviewers did comment that the person-based structure seemed to work better for the income questions, especially during callbacks to obtain missing data, and also that it offered some advantages in special circumstances, such as interviews with very elderly respondents, with "non-family" households, and in situations in which the respondent was using the mailed-out form as a response aid. (3.2.3)

Respondent Debriefing Questions The field experiment included a set of "respondent debriefing" questions administered after the completion of the main interview to assess respondents' reactions to the interview. In general, responses to these items indicate a preference for the topic-based interview. Compared to person-based respondents, topic-based respondents were more likely to report that they stayed interested throughout the interview, and were overwhelmingly less likely, when presented with the option, to express a preference for the other instrument structure. However, although the vast majority of respondents (over 95%) in both treatments reported that the interview "moved along smoothly," person-based respondents were significantly more likely to report this judgment. Also, those who experienced the person-based interview were more likely to agree that "everyone has a responsibility to answer surveys like this," and to disagree that such surveys are "a waste of people's time." These somewhat counter-intuitive results may be a result of effort justification, which social psychological research has shown to affect judgments of relatively more unpleasant tasks. They may also reflect a selection bias, if the topic-based instrument's lower refusal rate was associated with greater success in completing interviews with reluctant respondents. Other debriefing responses were found to be affected by the relatedness of household members. For example, person-based and topic-based respondents overall were about equal in their tendency to label the ACS/CATI questions as "repetitious" — however, in households in which all persons were related, person-based respondents were significantly more likely than topic-based respondents to apply the "repetitious" label; in non-related households the reverse was true. (3.3)

Behavior Coding We conducted a behavior coding analysis of approximately 200 interview cases, looking primarily for instances in which interviewers deviated in major ways from the interview script, and for evidence of respondents' difficulties in providing immediate and appropriate responses to the interview questions. This analysis indicates no significant differences between instrument treatments in the question-reading behavior of interviewers or the question-answering behavior of respondents. (3.4)

Item Nonresponse Differences by instrument type in missing data were assessed first with a global measure of the proportion of "on-path" items lacking a valid response (generally, a "DK" or "refused") for each interviewed person. This analysis shows no significant difference between the two instruments in the overall tendency to produce missing data. We also examined nonresponse

on an item-by-item basis. Here the results clearly favor the topic-based instrument. Twenty-nine items with "important" levels of nonresponse (generally, a nonresponse rate of at least 2%) showed significant nonresponse differences by instrument treatment — 24 of these differences favored the topic-based instrument, versus only 5 which favored the person-based instrument. (3.5.1)

Response Distributions Response distributions were analyzed for 38 ACS-CATI items, of which 11 (race, citizenship, current school attendance, speaking a language other than English, difficulty seeing/reading, working last week, riding to work with others, kind of employer, receipt of wage/salary income, self-employment, and receipt of interest/dividend income) show a significant difference between the two instruments. In general, the topic-based instrument seems to have elicited more reports of more rare characteristics — e.g., more Asian/Pacific Islanders, more naturalized citizens and non-citizens, more non-English speakers, etc. While intriguing, the implications of these differences for data quality are not clear. (3.5.2)

Within-Household Response Consistency An initial concern about the topic-based design was that it might encourage a tendency to overreport that all household members shared some characteristic. We analyzed six variables and found that for three (race, Hispanic origin, and current school attendance) there was no difference between the two instruments in the proportion of households in which all members shared the same characteristic. Three other variables did show such differences — the person-based instrument elicited more within-household uniformity regarding citizenship, while the topic-based instrument elicited more uniformity regarding "mobility" ("Did ... live in this house/apartment 5 years ago?"), and speaking a language other than English. Given the inconsistency in the direction of the effect, and the absence of validating data, it is impossible to draw from these findings any general conclusions about instrument differences in the tendency to elicit spurious uniformity, or spurious non-uniformity. (3.5.3)

2. Introduction/Background

2.1 The American Community Survey

The American Community Survey (ACS) is the monthly household survey program which is the data collection cornerstone of the Census Bureau's new "continuous measurement" (CM) system. CM is an alternative to the traditional, once-every-ten-years decennial census long form data collection. The primary goal of CM is to provide timely, annual updates of detailed housing, social, and economic data throughout each decade.

Following several years of testing and development, the ACS will be implemented in every county of the United States beginning in 2003. When fully operational, three million different addresses will be selected for sample each year. This will enable the ACS to provide estimates of the housing, social, and economic characteristics each year for all states, as well as for all cities, counties, metropolitan areas, and population groups of 65,000 persons or more. For smaller areas, it will take up to five years to collect data for the same number of households as are currently sampled via the decennial census long form. These multi-year estimates of characteristics will be updated each year

for every governmental unit, for components of the population, and for census tracts and block groups.

The ACS is conducted using three modes of data collection: self-enumeration through mail-out/mail-back methods; computer assisted telephone interviewing (CATI) for mail nonresponse cases for which a telephone number can be obtained; and computer assisted personal interviewing (CAPI) for a sample of mail nonresponding cases which cannot be completed by CATI. The research project described in this report evaluated an alternative design for the CATI instrument used in the middle stage of data collection.

More detailed information about the ACS, and about the CM program in general, is available at the Census Bureau's CM website: *www.census.gov/cms/www*.

2.2 Person-Based and Topic-Based Survey Instrument Structures

The research project that is the focus of this paper is an experimental evaluation of two different instrument designs for the ACS/CATI nonresponse followup survey phase, which we label the person-based approach and the topic-based approach. This section briefly describes the differences between the two designs.

In its initial formulations, the ACS/CATI survey instrument has followed a conventional person-based design for household survey questionnaires which are intended to gather data about all members of target households from a single household respondent. This design "decision" was in fact less a conscious decision than simply a direct translation of traditional paper-and-pencil questionnaire methods to the computer-assisted interview environment. The person-based approach in essence completes all topics for one person before proceeding to the next person, e.g.:

What is [person1]'s sex?
What is [person1]'s birth date?
What is [person1]'s marital status?
What is [person1]'s race?
Does [person1] have a work disability?
Has [person1] ever served in the Armed Forces?
[etc. for additional topics]

What is [person2]'s sex?
What is [person2]'s birth date?
What is [person2]'s marital status?
[etc.]

[repeat for persons 3, 4, etc.]

Before the advent of computer-assisted interviewing, the presence in the instrument of any even remotely complex branching patterns or skip instructions rendered the person-based structure the only practical design option. Computerization, however, has broadened the range of viable options to include a topic-based interview sequence. A topic-based interview completes one topic for all persons before proceeding to the next topic, e.g.:

What is [person1]'s sex?
What is [person2]'s sex?
[etc. for persons 3, 4,...]

What is [person1]'s birth date?
What is [person2]'s birth date?
[etc. for persons 3, 4, ...]

What is [person1]'s marital status?
What is [person2]'s marital status?
[etc. for persons 3, 4, ...]

[etc. for additional topics]

Moore (1996) summarizes the potential benefits (and a few potential pitfalls) of the topic-based approach, and provides some preliminary evidence from a small scale laboratory study which in general supports the notion that there are real, practical benefits to be gained from it. One of the key practical benefits of a topic-based interview derives from its ability to exploit the context established by the first presentation of the full text of a question to severely truncate the text needed for subsequent people, e.g.:

What is the highest degree or grade of school that [person1] has completed?
How about [person2]...?
And [person3]...?
[etc.]

In the next section we describe the development and early testing of prototype topic-based ACS instruments in a small-scale laboratory study. This pilot research not only provided sufficient positive evidence to justify a continued research effort, it also had a significant impact on the design of the topic-based instrument ultimately used in the large-scale main study.

2.3 Pilot Test Design and Results Summary

2.3.1 Background

In the late summer and early fall of 1996, we conducted a small-scale study of paper-and-pencil prototypes of the proposed topic-based ACS CATI instrument in the CSMR laboratory. The primary

goals of this study were to obtain insights into respondents' reactions to topic-based and person-based interviews, and to develop methods to assess those reactions and other evaluation criteria for use in the large-scale field experiment.

Initial plans for the laboratory study involved only two interview treatments. However, as we developed our topic-based interview materials for the laboratory study, it became apparent that the initial version of the topic-based instrument was not the optimal design. Specifically, through reviews that accompanied the development process, we became persuaded that the initial topic-based design was not sufficiently topic-based. In many cases where we had originally proposed grouping some related items within the topic-based design (e.g., age and marital status; Spanish origin, race, and their accompanying detail questions), we now felt that in almost all cases the interview would be improved to the extent that the "topics" in the topic-based design could be limited to individual items, not groups of items. We therefore added a third treatment group to the study, a more "extreme" topic-based instrument (T2), along with the original topic-based instrument (T1) and the person-based (P) control instrument.

2.3.2 Pilot test design

Using the Center's existing database of prospective respondents, augmented with names acquired through some special recruiting efforts, we recruited respondents and conducted 45 one-hour-plus laboratory sessions. Respondents were paid $30 for their participation in the study. Each session involved an ACS telephone interview (using one of the three interview treatments, P, T1, or T2) and a followup debriefing interview. Our two primary selection criteria for study respondents were that they be at least 18 years of age and members of relatively large (preferably 4 or more person) households. We subsequently discarded several interviews conducted with young adult children living with their parents who, although they met our age requirement, were very uninformed about employment, income, and other issues covered in the ACS interview. This resulted in a final effective study size of 37 cases — 15 with the person-based control instrument (P), 12 with the original topic-based test instrument (T1), and 10 with the revised topic-based instrument (T2).

The laboratory session consisted of a taped telephone interview and a followup debriefing session about the interview experience. The debriefing arrangements were modified somewhat after the initial interviews, but for most interviews the debriefing included both a self-administered questionnaire and an interviewer-administered interview. (Copies of the pilot test debriefing questionnaire and interview form are included as Technical Appendices 1 and 2.)

Four CSMR researchers conducted the ACS telephone interviews using paper-and-pencil prototypes of the planned CATI instruments. While one researcher conducted the telephone interview, another observed the respondent's behavior through a one-way mirror, noting particularly any overt evidence of dismay, fatigue, boredom, disengagement, etc.

2.3.3 Pilot test findings

In this section we briefly summarize the key results of the pilot test. Note that, because of the small and non-representative nature of the pool of respondents, we make no attempt to apply formal statistical tests to the findings.

i. Length of interview We expected that a topic-based interview would generate savings in interview length. As shown in Table 1, however, the pilot test results do not offer strong confirmation of this expectation. Overall, the more extreme (T2) topic-based interviews were substantially shorter to administer than the other interview treatments, but this difference appears much more modest when we control for the number of adults in the household (children under age 15 are skipped out of most ACS questions).

What was unexpected was the length of the T1 interviews, which appeared to be a step backward from the person-based approach in terms of interview efficiency. Recall, however, that all three treatments used paper-and-pencil prototypes which attempted to model CAI instruments. The researchers who conducted the pilot test interviews all found the topic-based questionnaires much more difficult to manage than the person-based forms, but especially the T1 instrument because of its additional paper-shuffling and thought-collecting demands. The test results may therefore represent the minimal gains likely to be experienced with a fully automated topic-based instrument.

ii. Respondent debriefing assessments Most of the debriefing questions posed to pilot test respondents (e.g., was the respondent frustrated during the interview, was the respondent bored, did the questions seem to come "out of the blue," did the respondent feel a desire to be done with the interview, etc.) showed virtually no difference in mean response among the three treatments. Three questions — about the felt "repetitiveness" of the interview questions, the level of impatience the respondent experienced, and the after-the-fact expressed preference for a different style of interview — did suggest important differences, as shown in Table 2.

The order in which the three instruments fall is the same across all three of the dimensions covered in Table 2, but it is really the T2 topic-based version that stands apart from the others. The T2 instrument was seen by respondents as the least repetitive, was the least likely to induce feelings of impatience, and elicited the lowest expressed preference for the alternative structure type.

iii. Respondent behavior observations We observed respondents' behavior through a one-way mirror during the pilot test interviews, looking especially for displays of any particular displeasure — or pleasure, although we expected such displays to be vanishingly rare (which, indeed, they were). We noted both verbal and non-verbal displays on coding sheets designed for this purpose (see Technical Appendix 3). Table 3 summarizes the key findings of these observations.

Despite the rough and subjective nature of the coding system, and the rather blurred distinction between categories of behavior, "displeasure" displays were clearly less frequent in T2 topic-based interviews than in the other interview conditions. Especially notable are the results for "boredom/fatigue" displays, which were about 40 times more common in the person-based and T1

topic-based interviews than they were in T2 topic-based interviews, where they were, in fact, quite rare.

2.3.4 Conclusions

To us, the pilot test results clearly suggested that the topic-based design was generally superior to the person-based design, and in particular that the T2 topic-based design was clearly an improvement over the original T1 design. The more extreme T2 format seemed to offer the potential for important reductions in interview time, and importantly reduced levels of impatience, boredom, and other negative affect in respondents. As a result, we recommended moving forward with a large-scale and more rigorous test of the topic-based approach, using a slightly modified T2-type topic-based instrument that almost exclusively equated "topic" with "individual question." These recommendations were accepted. In the next section we describe the design of the subsequent field test, the results of which are the primary focus of this report.

2.4 Field Test Design and Procedures

2.4.1 Schedule and workloads

The person-based/topic-based experiment was conducted at the Census Bureau's Jeffersonville Telephone Center (JTC) CATI facility. Interviewing was conducted from October 3 through 26 (for sample cases which had failed to respond by mail in September) and November 1 through 26 (for October mail nonrespondents). Cases were assigned at random to one or the other instrument type[2]. Table 4 shows, by month and for each instrument type, the number of addresses originally sent to the CATI operation (i.e., mail nonresponse cases for which a telephone lookup operation produced a telephone number), the number eventually found to be CATI-ineligible (for the reasons shown), and the resulting number of cases eligible for the CATI followup interview.

2.4.2 Staffing

Staffing arrangements were an important element in the design of the test. Our goal was to ensure to the maximum extent possible that the test of the two instruments was uncontaminated by differences in the skills or experience levels of the interviewers. We wanted generally experienced interviewers, but not ACS-experienced interviewers, who were available to work both months of the project — generally experienced, so that the "learning curve" up to standard productivity would be as brief as possible; not ACS-experienced, to avoid as much as possible any pre-set notions about the "proper" way to conduct the ACS interview; and available both months, in order to keep the interviewer pool as constant as possible throughout the course of the study.

---

[2]The mail-out questionnaires for this test included a split-panel experiment testing the impact of minor wording changes on income reporting. For purposes of the person-based/topic-based experiment these questionnaire differences can be ignored, since the instrument treatments were fully crossed with the paper questionnaire wording experiment.

For the most part these goals were met. The range of interviewing experience of the selected staff was from 2 months to 12 years, with an average tenure of about 19 months. Of the 44 interviewers who worked on the test in October, only two had prior ACS experience (with, of course, a person-based designed instrument); these two were included in the study because of their foreign language skills. There was some attrition after October, and a very small number of new staff were brought in to work on the test for the second month. In November, out of the staff of 38 interviewers used, six had prior experience with ACS (other than with the test), including again the two continuing October interviewers with special language skills.

JTC staff split the interviewers into two teams, with the goal of ensuring as much as possible that the teams were equivalent in skill levels and experience. Each team was assigned at random to either the person-based or the topic-based instrument for the first month of interviewing, and then switched to the other instrument for the second month.

2.4.3 Interviewer training

Two local supervisors — one for each instrument — conducted the initial and second month training sessions. The verbatim training packages for the two instruments were prepared by headquarters staff. Each initial package focused primarily on the ACS questions themselves, with some attention also devoted to the particular structure of the instrument that the interviewers were assigned to in the first month. The first month's training session was about 15 hours in length. In addition to the classroom training, interviewers were given a two-week period before the start of "live" interviewing in October in which a training sample of cases was made available to them for practice interviews. Before the second month's interviewing, the interviewers received an additional 5 hours of training on their new instrument. Because the interview content was unchanged, the second training focused almost exclusively on the differences in the way the second month's interviews would be structured.

It is perhaps worth noting that in at least one respect, the training interviewers received was not exactly equivalent across the two instruments. In order to avoid what would be an extremely stilted and unnatural-sounding interaction if the interview script as presented were to be followed precisely, the topic-based training explicitly permitted interviewers, at their discretion, to use first names only when referring to household members in sequence (e.g., "What is John's date of birth? How about Mary? And James? And how about Susan?"). This same leeway was not offered in the training for person-based interviewing. We suspect that interviewers conducting person-based interviews often took similar liberties, regardless of their training instructions, but in fact we have no evidence concerning any differences by instrument type in interviewers' treatment of names during the ACS/CATI interviews.

3. Results of the Person/Topic Experiment

In this section we summarize the major results of the experiment. We organize our results in five sub-sections: (1) interview outcomes, in which we present analyses of cooperation rate and interview length differences between the two instruments; (2) staff assessments, in which we

summarize the results of various efforts to gather interviewers' (and supervisors') judgments concerning the comparative strengths and weaknesses of the two instrument designs; (3) respondents' assessments of the interview experience as a function of type of instrument; (4) the nature of the interviewer/respondent interaction, as indicated by behavior coding; and (5) data outcomes, in which we examine item nonresponse differences, differences in the response distributions for selected items, and differences between the two instruments in the tendency to produce within-household consistency on a small set of characteristics.

3.1 Interview Outcomes

3.1.1 Household response/refusal rates

The results of the person/topic test indicate a small but statistically significant advantage for the topic-based instrument in the completion of interviews among assigned, eligible cases, as well as a reduction in the proportion of cases not completed due to refusals to participate. Table 5 summarizes household response outcomes by instrument type for the two months of the person/ topic test. For the two months combined, summarized in the last two columns of Table 5, a simple t-test indicates significant differences by instrument type in both the completion rate and the refusal rate ($t=2.53$ and 2.56, respectively; $p<.05$, two-tailed).

We note that the test results may minimize the true difference in completion rate between the two instruments. For both instrument treatments in October, and for the topic-based instrument in November, between 2.6% and 3.2% of the completed interviews were completed by high-level supervisory staff. These rates seem reasonable, in both level and consistency, given agreed-upon procedures — namely, that supervisors' interviewing would be restricted to handling respondent-initiated incoming calls and occasional refusal conversion. However, the rate of supervisor-completed interviews was about three times higher than usual for the person-based treatment in November, accounting for 8.7% of all completes. In this case, one supervisor, reportedly very skilled and experienced as an interviewer, did substantial interviewing work in November on the person-based instrument. The effect of this lack of balance was to award some advantage to the person-based treatment group, with perhaps a dampening effect on the true difference in completion rates between the two instrument types.

The significant refusal rate advantage enjoyed by the topic-based instrument presents an interesting logical puzzle. As is typically the case (Groves and Couper, 1998), mid-interview "breakoffs" in this study were very rare; virtually all refusals occurred during pre-interview "negotiations," well before the structure of the interview was even potentially apparent to the refusers. Thus, the refusal rate difference between the instrument treatments in this test seems most likely to have arisen from differences in interviewers' behaviors in the face of similar base rates of respondent reluctance to participate. One possibility is that the difference in refusal rate is another facet of interviewers' preference for the topic-based instrument, here manifested in a reluctance to put forth quite as much effort to initiate a potentially difficult interview when the instrument to be used for that interview was of the less favored person-based variety.

3.1.2 Interview length[3]

Table 6 shows that the topic-based instrument resulted in a significantly shorter average interview for all household sizes up to and including 5 persons. Although by observation that difference is reversed in households containing 6 or more people, statistical testing fails to find a significant difference among the largest households. Across all household sizes the average topic-based interview was about 2 minutes shorter than the average person-based interview, a difference that is both statistically significant and, given the eventual scale of the ACS operation, economically significant as well. Moore (1996) estimates that reducing the average CATI interview length by this amount in the full-production ACS environment would yield annual savings of approximately $300,000 in interviewer labor costs alone.

3.2 Staff Assessments

We offered the telephone facility staff several opportunities to assess the instruments during the two months of testing. Interviewers completed a debriefing questionnaire at the end of each interview month. In addition, nine interviewers took part in a debriefing focus group held near the end of the interviewing in the second month. By that point the interviewers had substantial experience with both instruments and could provide comparisons and a design preference if they had one. After the first month of interviewing, supervisors also completed a debriefing questionnaire that solicited their comments on the two instruments. The results from these staff assessments are summarized below.

[3]We thank Gregg Diffendal (CMS) for producing the interview length results and for carrying out the analysis of them.

3.2.1 Supervisors' debriefing questionnaires

As is typically the case, supervisors carried out a wide variety of tasks in the ACS person/topic test. Most importantly, they provided survey informational and technical support for interviewers. As noted above, two supervisors trained the interviewers. The supervisors were also called on occasionally to conduct CATI interviews — in particular, when respondents called in and initiated an interview, and also in "refusal conversion" cases in which a respondent's reluctance to participate could not be overcome by the interviewer. A total of 32 supervisors worked at least some time on the ACS test over its two-month duration, although the number of supervisors who devoted significant time to the test was considerably lower than that.

In early November, supervisors were provided with a debriefing questionnaire requesting their assessments of various aspects of the two instruments. Unfortunately, only six supervisors returned completed debriefing questionnaires. Below we summarize the key points from the supervisors' debriefing responses. An important caveat to these conclusions, of course, is the very real risk that the limited response to the debriefing questionnaires renders them not representative of the opinions of the entire group of supervisors.

In general, the supervisors' debriefing responses do not reveal a clear preference for one or the other instrument. Each instrument format generated both positive and negative comments, although it should be noted that some of the responding supervisors admitted to a bias in favor of the person-based approach because of their experience — in some cases quite extensive experience — with that design from their prior ACS work. Among the noted positive features of the person-based instrument were the following:

- ease of training, since it was so similar to the other ACS training they had done;
- the greater formality of the person-based interview (some of the supervisors were clearly uncomfortable with the use of first names during the interview, which the topic-based instrument explicitly allowed, because it made the interview "too friendly — less professional");
- reduced awkwardness in situations involving poorly-informed proxies (although in fact supervisors expressed a desire that both instruments be designed with the capability to skip over, and call back later, a person whose data could not be provided by a proxy[4]);
- the greater ease of concentrating on, and keeping track of, one person's data; and
- its more general appropriateness — one respondent commented that the person-based form "works well for ALL situations."

[4]Callbacks to complete partial interviews comprised the bulk of the interviewing work that fell to supervisors, so it is hardly surprising that their remarks focused on these overall rather rare situations in which the interviews were most awkward — that is, when proxy data collection caused difficulty in the progress of the interview.

All comments regarding positive aspects of the topic-based instrument focused on perceived reductions in interview times. Other positive comments noted that the topic-based design "probably resulted in much happier respondents," and also that it was "not as repetitious as person[-based] for large households."

3.2.2 Interviewer debriefing questionnaires

Interviewers were provided three formal opportunities to evaluate the test instruments: two debriefing questionnaires and a debriefing session/focus group. (The focus group and its findings are described in section 3.2.3 below.) The debriefing questionnaires were distributed about two weeks into each survey month. The two administrations enabled us to obtain both the "naive" opinions of the interviewers about each instrument (i.e., before they had any experience with the other instrument), and their opinions at the end of the experiment, when they could directly compare the alternate designs.

Only 36 of the 44 October interviewers completed a debriefing questionnaire in October; there was also some nonresponse in November, when only 32 of the 38 interviewers completed a debriefing questionnaire. In addition to this debriefing form nonresponse, interviewers sometimes failed to provide responses to individual debriefing questions. In the tables that follow, we show in each instance the number of responding interviewers, but we make no attempt to adjust for nonresponse. In addition, because of the nature of the data and the small number of cases, we treat these results more as impressionistic observations than as statistical data to be analyzed with statistical tests.

i. Interviewers' likes and dislikes for each instrument version The debriefing questionnaires included open-ended questions asking interviewers what they liked about the instrument assigned to them that month and what they disliked about that instrument. As Table 7 shows, interviewers had many more positive things to say about the topic-based instrument than the person-based instrument, and many more negative things to say about the person-based instrument than the topic-based instrument.

The differences by instrument type were apparent in the first month of the test, before interviewers had any experience with the "other" design. In both months, interviewers were about twice as likely to offer a "like" comment to the topic-based instrument than to the person-based instrument, and those who offered "like" comments to the topic-based instrument provided about 40-50% more "like" comments per commenter than did those responding to the person-based instrument. The results in Part B, the "dislike" half of the table, present almost a perfect mirror image to the "like" results.

The main reason offered for liking the topic-based instrument was that it made for a faster interview. Interviewers also simply liked the technique of asking questions in the topic-based format — i.e., asking a question of everyone before proceeding to the next question — and felt that the ability to use the abbreviated "prompts" reduced their effort substantially. The most common reasons offered for liking the person-based instrument had to do with its greater structure and "orderliness." For both instruments, however, the foci of the "dislike" responses were far more concentrated. The

person-based instrument was seen as too long, and highly repetitive and tedious; the topic-based instrument was faulted for not working as well in roommate/boarder households and other situations requiring callbacks to complete the interview.

ii. Perceived design advantages The debriefing questionnaires asked interviewers to rate the extent to which their assigned instrument contributed to an improved interview on five different dimensions. Each of these items used a 5-point rating scale ranging from "strongly agree" to "strongly disagree." Table 8 shows the proportion of positive ("strongly agree" plus "agree") responses for each of the five dimensions.

By generally overwhelming margins, the interviewers favored the topic-based instrument as the one that made the interview flow more quickly and naturally, helped them become familiar with the organization of the questionnaire, made it easier for them to conduct the interview, and made it easier for them to probe for more correct answers. These differences are quite apparent in the October results, before the interviewers had a chance to compare the two instrument designs directly, but seem to have become even more marked in November, at which point both groups of interviewers had experienced both instrument versions.

iii. Perceived performance with different types of households Interviewers were asked to rate the performance of their assigned instrument design in various types of households. Again, we used 5-point scales ranging from "very well" to "very poorly." Table 9 shows the proportion of interviewers who rated their instrument positively (i.e., who gave a response of "very well" or "well") for each type of household.

These results clearly indicate interviewers' general preference for the topic-based instrument in all types of households — with the exception of single-person households, where the interviewers appear to have given a slight advantage to the person-based design[5]. The results for households with unrelated persons and elderly households are especially interesting. In both cases, contrary to the comments interviewers made during the debriefing session (see section 3.2.3 below), interviewers' debriefing questionnaire responses suggest that they perceived the topic-based instrument to have performed better than the person-based version.

iv. Anticipated and actual instrument preference The debriefing questionnaire for each month asked interviewers to choose a preferred instrument format. In the October questionnaire the question was a hypothetical one[6]; in November the interviewers were simply asked their preference directly (see the question texts in Table 10, below). In both months, regardless of question wording

[5]The suggestion of any difference between the instruments in one-person households is something of a mystery, since the two designs were absolutely indistinguishable in such circumstances.

[6]The question posed to interviewers in the interviewer debriefing questionnaire in October was virtually identical to one included in the respondent debriefing questions at the end of the ACS interview in both months; see section 3.3, below.

or interviewers' experience, the overwhelming preference of the interviewers was for the topic-based instrument, as shown in Table 10.

3.2.3 Interviewer debriefing focus group

In addition to the interviewer debriefing questionnaires, CM staff conducted a debriefing focus group session with nine interviewers in mid-November, near the end of the second month's interviewing. The group included five interviewers who had begun with the person-based instrument and then switched to the topic-based, and four who had experienced the instruments in the reverse order. The session lasted about two hours. (Technical Appendix 4 is a copy of the moderator's guide used in conducting the debriefing focus group.)

Not surprisingly, given their debriefing questionnaire responses, the interviewers expressed many more positive comments about the topic-based instrument than they did regarding the person-based instrument. Positive comments about the person-based instrument were not entirely absent, however. Echoing the comments of the supervisors on their debriefing questionnaires, the interviewers were generally of the opinion that callback interviews to obtain data for selected individuals were easier with the person-based design. Some interviewers commented that income reporting seemed easier to manage with the person-based instrument. Interviewers reported that in some cases respondents wanted to use their not-returned paper questionnaires as a guide during the telephone interview; the person-based approach was also felt to work better in this situation, since it follows the paper questionnaire design.

In general, however, the focus group participants clearly preferred the topic-based instrument. It seemed faster to the interviewers (although some expressed doubt that it actually <u>was</u> faster). They felt they spent less time talking and this made the job easier for them. One interviewer noted a respondent's comment that she was glad she hadn't completed the paper questionnaire because the topic-based interview seemed shorter to her than filling out the form would have been. Another interviewer reported that she felt she obtained more honest answers with the topic-based instrument because respondents just answered the questions for all household members and didn't pause repeatedly — as they had the opportunity to do with the person-based design — to evaluate how truthful or committed to completing the interview they wanted to be. The interviewers as a whole felt that the topic-based was a more conversational instrument, which enabled them to establish rapport more easily and to conduct a more relaxed interview. Unlike the supervisors, the interviewers liked being able to use respondents' first names, which they felt reduced an artificial barrier in the interview. Interviewers also expressed the feeling that respondents had more confidence in them with the topic-based approach, because having to read repeatedly the full text of each question in person-based interviews revealed their role as a mere script reader, as opposed to someone seeking information in a more natural, conversational manner.

3.3 Respondent Assessments — Respondent Debriefing Questions

The telephone interviews included a set of "respondent debriefing" questions, administered after the completion of the main interview, the purpose of which was to assess respondents' reactions to the interview. In this section we describe the results of these debriefing questions.

3.3.1 Respondent debriefing analysis details

i. Analysis exclusions The analysis of the debriefing responses excludes certain cases, as outlined in Table 11 and the text which follows:

No permanent residents: About 1% (n=28) of the 2266 completed interviews were conducted in households with no permanent residents. The ACS system classifies these as vacant units, so the interview in these cases includes only the housing questions. Because of their special and very truncated interview, for which the person/ topic distinction is irrelevant, these cases are excluded from the respondent debriefing analyses.

One-person households: Person/topic instrument design differences are irrelevant in one-person households; because such households are not informative regarding the different impact of the two designs on respondents, the cases used for analysis include only households containing two or more people. One-person households comprised 502 of the 2238 interviews completed in occupied residences — 243 in the person-based group, and 259 in the topic-based group, in each case about 22% of all occupied/complete interviews.

Non-continuous interviews: The debriefing questions were only administered to respondents who completed the ACS interview in one uninterrupted session — i.e., they were skipped in interviews which required a call-back in order to be completed, or in which a second person served as a respondent for part of the interview. One of these conditions occurred in 236 of the interviews completed in 2-or-more-person households. Unlike the other exclusions, however, this one was clearly not equivalent across interview treatments: (141/855=) 16.5% of the otherwise eligible person-based interviews were missing all debriefing data, compared to only (95/881=) 10.8% in the topic-based group. Excluding the "debriefing data missing" cases yields a final analysis sample of 714 person-based cases and 786 topic-based cases, for a total of 1500 cases.

ii. Limitations As noted in section 3.1.1, the person-based treatment experienced a significantly higher interview refusal rate than the topic-based treatment. One likely result of this differential nonresponse is that the person-based debriefing responses might be positively biased (relative to the topic-based treatment), due to the greater "weeding out" of more disgruntled (non)respondents at the beginning of the interview. The person-based treatment also experienced more attrition from the debriefing questions due to interview "interruption" (see above), which may also have affected the comparability of the two treatment groups, although with unknown implications for the debriefing results. The analyses presented below ignore both the exclusion of whole cases from the debriefing questions and item nonresponse among those for whom the

debriefing questions were administered. Although for most debriefing items the number of "don't know's" and "refusals" is quite small, for others the level of nonresponse is sufficient to raise some concerns about possible nonresponse biases. (Note also that this report makes no attempt to distinguish the two forms of item nonresponse.)

iii. Analysis procedures For the primary focus of the experiment — differences in the debriefing responses according to instrument design — statistical significance tests were carried out using SAS-PC's "proc ttest" procedure. For dichotomous variables, the t-test procedure essentially assesses the difference between two proportions. For non-dichotomous variables, the t-test procedure assesses the difference between treatment means assuming a 3- or 4-point scale (as indicated in the tables below). In the tables, significance levels are reported for a two-tailed test. "Scaled" variables were also analyzed with chi-square tests.

A secondary focus is the effect on interview outcomes of the presence in the household of non-related individuals, which is hypothesized to present greater difficulties for a topic-based interview than for a person-based interview. We define non-related households as those containing at least one person whose relationship to the reference person is "roomer/boarder," "housemate/roommate," or "other non-relative," which we find to characterize (52/714=) 7.3% and (53/786=) 6.7% of the person-based and topic-based cases, respectively. We test for the moderating effects of non-relatedness (on instrument-design-based differences in debriefing responses) via a modeling exercise, using SAS's "proc glm" procedure to assess the significance of an interaction between instrument type and the relatedness of household members.

3.3.2 Summary of respondent debriefing results

i. Main effects of instrument design Responses to the debriefing questions suggest several ways in which those interviewed with the person-based instrument and those interviewed with the topic-based instrument experienced the ACS/CATI followup interview differently. In general, and despite the possible biasing effects of differential survey nonresponse to the contrary, respondents' replies to the debriefing items suggest a preference for the topic-based instrument. Compared to person-based respondents, topic-based respondents were more likely to report that they stayed interested throughout the interview, and overwhelmingly less likely, when presented with the option, to express a preference for the other instrument structure. Person-based respondents, however, were more likely to report that they felt the interview "moved along smoothly." Those who experienced the person-based interview were more likely to agree that "everyone has a responsibility to answer surveys like this," and to disagree that such surveys are "a waste of people's time." We offer as possible explanations for these somewhat counter-intuitive results both selection bias due to differential nonresponse, and the notion of effort justification from the field of social psychology.

ii. Effects of "non-relatedness" Debriefing responses were found to be affected by the relatedness of household members in only two instances. First, while person-based and topic-based respondents overall were about equal in their tendency to label the ACS/CATI questions as "repetitious" (see above), this apparent similarity masked substantial differences between related and

non-related households. In households in which all persons were related, person-based respondents were more likely than topic-based respondents to apply the "repetitious" label; in non-related households the reverse was true. The second effect of the relatedness of household members is evident in responses to the debriefing item which described the other possible design of the ACS/CATI instrument and asked respondents whether they might have preferred that design to the one actually used in their interview. In all-related households, person-based respondents were much more likely than topic-based respondents to express dissatisfaction with their interview by endorsing the "other" instrument design; this difference was significantly reduced (though not reversed) in non-related households.

3.3.3 Detailed results

i. Perceived appropriateness of interview length Regardless of whether one looks at the entire "too short/about right/too long" scale (Table 12), or a recoded version collapsing the very sparse "too short" category into a combined category with "about right" (data not shown), there is no evidence of any difference between person-based respondents and topic-based respondents in how they perceived the length of the ACS/CATI followup interview. In each instrument group, about 55% of the respondents judged the length of the interview to be "about right," while a little less than 45% said that it was "too long." Only a very small handful of respondents reported that the interview was "too short." The presence or absence of non-related household members had no significant effect on these results (analysis not shown).

ii. Felt boredom/impatience Respondents who experienced the topic-based interview were significantly more likely to report that they "stayed interested" throughout the interview; correspondingly, person-based respondents were more likely to report getting "bored or impatient" (Table 13a). Combining this item with its followup item to form a scale of felt boredom/impatience yields a somewhat mixed picture. A chi-square test indicates that the two treatment groups differed in the distributions of their responses on the resulting 3-point scale, although a t-test on the difference between the treatment means is not significant (Table 13b). What this suggests is that topic-based respondents, while less likely to report that they were bored or impatient, were more likely, if they did report being bored, to place themselves in the more extreme boredom category. Statistical analysis (not shown) indicates no significant impact of the presence of non-related household members on the greater tendency of person-based respondents to report getting "bored or impatient."

iii. Perceived interview "flow" The overwhelming majority of respondents in both instrument treatments reported a positive impression of the "flow" of the interview. In each case, 95% or more of the respondents reported that they felt that the interview "moved along smoothly," and no more than 5% felt that it sometimes seemed to "jump around" (Table 14). Although the absolute difference between the groups was small, person-based respondents were significantly more likely to endorse the "moved along smoothly" position than were topic-based respondents. Additional analysis (not shown) indicates that this difference was consistent across both related and non-related households.

iv. Perceived repetitiveness of the interview questions Overall, the two instrument treatment groups did not differ significantly in their responses to a single debriefing item about whether the interview questions seemed "repetitious" (Table 15a). However, this surface similarity masks a substantial difference which becomes apparent when the relatedness of household members is taken into account (Table 15b). In households in which all persons were related, person-based respondents were more likely than topic-based respondents to apply the "repetitious" label; in non-related households the reverse was true.

The "relatedness" results in Table 15b are largely mirrored in a scale combining the "repetitious" item and a followup assessing how repetitious the questions were perceived to be. In related households, person-based respondents tended to place the ACS interview questions more toward the "very repetitious" end of the scale than did topic-based respondents; in non-related households, topic-based respondents viewed the questions as more repetitious (Table 15c).

v. Preference for a different interview structure When their own interview structure (person-based or topic-based) was described and contrasted with the other possible structure, respondents who experienced the person-based interview were about three times more likely than topic-based respondents to endorse the other instrument structure as a way to produce a "smoother" interview (Table 16a). Not surprisingly, interviewers did not particularly like this debriefing question — it IS a major mouthful — and reported in the interviewer debriefing session that they often had to paraphrase it in order to enable respondents to understand it. Note that, despite these efforts, the missing data frequencies for this item are still higher than for most other items. Although the effects of elevated nonresponse and of interviewers' special efforts are uncertain, they represent an important caveat in the interpretation of this debriefing item.

The more detailed analysis examining the impact of the non-relatedness of household members suggests that this characteristic did significantly affect responses to the instrument preference item (Table 16b). In this case, the highly significant difference (to the advantage of the topic-based instrument) in all-related households is reduced substantially — but, it should be emphasized, certainly not reversed — in households containing non-related persons.

vi. Responsibility to cooperate with "surveys like this" Compared to topic-based respondents, respondents who experienced the person-based interview were significantly more likely to agree that "everyone has a responsibility to answer surveys like this" (Table 17a). This somewhat counter-intuitive finding (in the context of other positive results for the topic-based instrument) may be a manifestation of a selection bias due to differential nonresponse (see sections 3.1.1 and 3.3.1.ii). It is also consistent with social psychological research on effort justification (McGuire, 1968). In this case, those who cooperated with a relatively more unpleasant interview (the person-based interview) seem to have felt more compelled to identify a justification for their behavior (a greater perceived duty to respond) than did those for whom the interview experience was less onerous.

The results for the expanded 4-point scale combining the "responsibility" item and its followup are consistent with the results for the individual item (see Table 17b). The significant chi-square

indicates that the response distributions for the two treatment groups differed significantly, and the t-test results confirm that the mean scale score for person-based respondents was significantly closer to the "strongly agree" (that there is a general responsibility for people to answer surveys "like this") end of the scale. Additional analysis (not shown) indicates no significant impact of the presence of non-related household members on responses to the "responsibility" items.

vii. Perceptions that surveys "like this" are a waste of time Compared to topic-based respondents, respondents who experienced the person-based interview were significantly less likely to agree that "surveys like this one are a waste of people's time" (Table 18a). As with the previous "responsibility to answer" results, this difference may reflect a selectivity bias, or it may reflect some cognitive work to justify the extra effort required to complete the less desirable interview.

The results for the 4-point scale combining the individual agree/disagree item with its followup are generally consistent with the results for the individual item (see Table 18b), although a chi-square test does not find a significant difference in the distributions of the responses for the two instrument treatments, and the difference between the means is only marginally significant. Additional analysis (not shown) indicates no significant effect of the presence of non-related household members on responses to the "waste of time" items.

3.4 Interviewer/Respondent Interactions — Behavior Coding

Behavior coding has come to be commonly used as a tool for evaluating survey questions, and in particular for identifying questions which cause problems for interviewers to read, or for respondents to answer, or both (Fowler and Cannell, 1996). We implemented a behavior coding analysis not to look at individual questions but to compare the person-based and topic-based instruments more globally — to determine whether they differed overall in their effects on interviewers' behaviors in asking the interview questions or respondents' behaviors in answering them. In this section we briefly describe the design of the behavior coding effort and its results. Overall, we find very little evidence of any difference between the two instruments in the frequency with which interviewers experienced difficulty in administering the questions or the frequency with which respondents experienced difficulty in providing adequate answers.

3.4.1 Design/methods

In order to yield sufficient cases for analysis, we established procedures which would result in approximately 50 tape-recorded interviews per instrument treatment per month, or about 200 cases altogether. We did not attempt to rigorously sample cases for inclusion in the analysis; each month, after allowing interviewers a brief "warm-up" period to achieve some familiarity and comfort with the assigned instrument (generally, a day or two of interviewing), supervisory staff simply instructed each interviewer that the next several interviews were eligible for tape recording. Scripted into each instrument were an eligible-for-taping (yes/no) screen and the necessary permission requests and taping explanations for to-be-taped cases. As noted above, the targeted number of completed interviews to tape record was 50 per month per instrument, excluding interviews in one-person

households and those which were not completed in a single interview session. Difficulties with the trace files (the computer version of the interview) and miscellaneous other technical problems reduced the actual number of interviews available for analysis to 182 cases, as shown in Table 19, below:

The behavior coding scheme used for this study was quite standard, both in its primary focus on the first interviewer-respondent "exchange" generated by each question, and in the array of behaviors (and their definitions) of primary interest. (Technical Appendix 5 offers more complete details on the coding scheme and its implementation.) The only notable addition to the coding task was the coder's assessment, for each coded question, of whether the interviewer's entry for the item matched what the respondent actually said. Experienced, trained behavior coders at the Census Bureau's Hagerstown (MD) Telephone Center carried out the coding using the "trace files" and the audiotapes from the interviews simultaneously. The trace files essentially produce a reenactment of the interview on a PC monitor, thus presenting the coders with the appropriate question script and response task against which to assess interviewers' and respondents' behaviors, and also freeing the coders from having to retrace the path of the interview on their own. The codes were entered on coding sheets, which were subsequently keyed by CSMR support staff.

3.4.2 Results

The primary focus of the behavior coding evaluation was to compare the two instruments as a whole, rather than to identify problematic individual questions. Therefore, for each interview we tallied the number of items coded to each interviewer behavior and each respondent behavior; dividing each tally by the total number of "on-path" questions (and then multiplying by 100) yields a percent score for each behavior coding category for each interview. Table 20 presents the average of all the scores across each instrument type. (For reference, Table 20 also shows the number of interviews on which the behavior coding results are based, and provides the range and average of the number of "on-path" questions per interview.) Table 20 shows that person-based interviewers used the exact (or only slightly changed) question wording an average of 72.5% of the time in administering questions to respondents; the comparable figure for topic-based interviewers was a very similar 70.7%. On the respondent behavior side, person-based respondents waited for the interviewer to complete the question and then provided an adequate answer an average of 76.0% of the time; for topic-based respondents the average percent of complete question/adequate answers was nearly identical at 75.6%.

Rough statistical tests (not shown) readily confirm what simple observation suggests — that there are no important differences between the two instruments on either interviewer or respondent behaviors. Person-based and topic-based interviewers held to their respective interview scripts at approximately equal rates, just as respondents provided adequate answers at about the same rate regardless of the type of instrument.

As noted earlier, coders also indicated their assessment of whether the interviewer correctly entered the response supplied by the respondent. While some baseline level of keying errors on the part of

interviewers is inevitable, we suspect that additional such response/entry mismatches are likely to the extent that a survey's questions fail to elicit immediately adequate responses from respondents, and that respondents' answers only emerge in the "noisier" give-and-take conversation that interviewers and respondents must occasionally engage in. Table 21 shows that, although the proportion of confirmed matches is somewhat disturbingly low, there is no important difference between the two instrument versions in the quality of the survey data, at least according to this rough indicator. This finding is consistent with the general results of the behavior coding shown above, which suggest that the frequency of extra "conversations" about questions and answers was about the same for the two instruments.

Additional analyses of the effects of household size on interviewers' and respondents' behaviors, summarized in Table 22, reveal a general decline in interviewers' exact (or near-exact) use of the interview script with increasing household size, but no obvious trend in respondents' tendency to supply an adequate answer (or in the match between reported and entered data; data not shown). However, most importantly for purposes of the current research, these general results appear to be entirely unaffected by instrument type.

3.4.3 Summary of behavior coding results

The behavior coding data suggest that the quality of the interview interaction did not differ in important ways according to the type of survey instrument used.

3.5 Data Outcomes

The content of the responses produced (or, in the case of item nonresponse, not produced) by the two ACS instruments can also be used to assess the impact of instrument design. In this section we examine three such data impacts. The first, item nonresponse, presents the most unambiguous evidence concerning the differential effects of the person-based and topic-based instruments on data quality — in fact, in the absence of validating data, it is the only available unambiguous evidence. The other comparisons examine instrument differences in the response profiles for a large set of survey items, and, for a much smaller set of items, the tendency of each instrument to produce consistent reports for all household members on some characteristic (e.g., race, language spoken at home, etc.)

3.5.1 Item nonresponse

We used two different techniques to assess instrument-based effects on item nonresponse. Both analyses exclude one-person households, because, as noted above, the two instruments produce identical interviews in such households, and inclusion of one-person households would merely dilute any real differences between different interview styles.

First, we calculated a "response completeness" index for each interviewed person — essentially, the proportion of "on path" items for which a non-missing response was entered — and then averaged these indices across all interviewed persons for each instrument type. (Note that we treat all interviewed persons as if their data were derived independently, when in fact they were not. In virtually all cases, a single household respondent provided data for him/herself and all other household members.) These results are presented in Table 23. Clearly, there is no difference between the two instruments in the extent to which they elicited complete interview data for all interviewed persons[7].

The second type of nonresponse analysis was a direct, item-by-item comparison of missing data rates across the two instruments. In Table 24, below, we show item nonresponse rates by instrument type for all items for which either instrument's nonresponse rate reached or exceeded 2%. (In other words, we ignore nonresponse rate differences where the overall level of nonresponse is trivial[8].) Where there is a significant difference (we use .10 as the cutoff for statistical significance), the higher rate is shown in double underlined font; n's shown indicate the total number of "on-path" cases, for which a response was expected.

Unlike the more global missing data indicator, the results of the individual item nonresponse analysis clearly indicate an advantage to the topic-based instrument (assuming that completeness of response is equally important across all content areas). Table 24 shows 29 significant nonresponse differences among items with "important" levels of nonresponse. For only 5 of those 29 differences is the nonresponse rate lower for the person-based instrument, versus 24 differences in which nonresponse is significantly lower for the topic-based instrument. We have no ready hypothesis to explain the apparent discrepancy between these results, which so clearly suggest better item nonresponse performance for the topic-based instrument, and the global "response completeness" results, which do not.

[7]Forming the index required special rules when the instrument path was indeterminate due to missing entries. When the instrument path could not be determined, the "response completeness" index assigned a path which maximized the number of questions in the path. Again, we thank Gregg Diffendal (CMO) for producing the global item nonresponse results and for carrying out the analysis of them.

[8]For the record, seven items with "trivial" levels of nonresponse nevertheless show a significant difference between instrument treatments. The differences are evenly split; in three cases the person-based instrument's nonresponse rate is lower than the topic-based, and in four cases it is higher.

3.5.2 Item response distributions

We compared the response distributions for the two instruments for 38 ACS/CATI items and found 11 items for which the distributions differed significantly. In this section we show the results for the significant comparisons[9]. (Our analyses generally ignore nonresponse; however, in the tables below we note all cases where nonresponse exceeds 2%, and thus could potentially affect the results obtained.) In general, the topic-based instrument seems to have elicited more reports of more rare characteristics — e.g., more Asian/Pacific Islanders, more naturalized citizens and non-citizens, more non-English speakers, etc.

As noted above, however, these measurement differences, while intriguing, provide no real clues as to any measurement *quality* differences between the two instruments. This issue clearly requires additional research. We also note that while the number of observed significant differences (11) is greater than what would be expected due to chance (4, using a .10 cutoff for statistical significance), several of the effects seem likely to be manifestations of a single underlying phenomenon, or are at least highly associated — e.g., more Asian/ Pacific Islanders, more naturalized citizens and non-citizens, and more non-English speakers.

i. Race As shown in Table 25, the significant difference between the person-based and topic-based response distributions for the ACS race item appears to be driven mostly by the difference in the A/PI category, where the topic-based rate of A/PI reports was about twice that of the person-based instrument.

ii. Citizenship Compared to the person-based instrument, the topic-based instrument elicited "native-born" reports at a lower rate, and other categories — especially naturalized citizens and non-citizens — at a higher rate. See Table 26. As noted, the results for this item may reflect the same underlying phenomenon that caused a significant difference for the race item (Table 25, above).

iii. Recent school enrollment Table 27 shows that the topic-based instrument elicited recent school attendance reports at a significantly lower rate than the person-based instrument.

iv. "At home" use of a language other than English Compared to the person-based instrument, the topic-based instrument elicited a higher reported rate of the use of a language other than English at home. See Table 28. Once again, differences for race, citizenship, and language may be separate manifestations of the same underlying process.

v. Reading disability Compared to the person-based instrument, the topic-based instrument elicited a higher rate of reported "reading disabilities," although the effect is only marginally significant.

[9]Attachment 1 lists the items for which the response distribution analysis showed no significant difference by instrument type.

vi. Worked last week Compared to the person-based instrument, the topic-based instrument elicited a higher rate of "worked last week" reports, although the effect is only marginally significant, as shown in Table 30.

vii. Rode to work with how many people? As shown in Table 31, compared to the person-based instrument, the topic-based instrument elicited a lower rate of "rode to work alone" reports, and a compensating higher rate of reports of riding to work with one other person.

viii. Category of employer: The (marginally) significant difference between the response distributions for this item appears to be driven mostly by the self-employed category; the rate of "self-employed" reports elicited by the topic-based instrument is about 30% higher than the rate for the person-based instrument. A difference in reported government employment may also be contributing to the overall effect. (Note that nonresponse for this item exceeds our arbitrary 2% cutoff level.)

ix. Receipt of any wage/salary income Compared to the person-based instrument, the topic-based instrument elicited a lower rate of reported wage/salary income receipt, although the effect is only marginally significant, and the missing data rate is very high. See Table 33.

x. Receipt of any self-employment income Compared to the person-based instrument, the topic-based instrument elicited a higher rate of reported receipt of self-employment income (see Table 34).

xi. Receipt of any interest/dividend income As shown in Table 35, the topic-based instrument elicited a rate of reported receipt of interest/dividend income that was about 25% higher than the rate of receipt reported by person-based respondents. Again, however, we note the presence of substantial nonresponse.

3.5.3 Within-household response consistency

The final data outcome that we examined in evaluating the performance of the two instruments was the propensity of each instrument to produce within-household consistency with regard to certain individual-level characteristics. Some ACS analysts were concerned that the topic-based format might encourage the over-consistent reporting. Presumably, either because the structure of the topic-based interview makes it easier for respondents to report about (and for interviewers to record information given about) all household members simultaneously, or because it permits interviewers to slide into a household style of questioning ("Is anyone here of Hispanic origin?"), the topic-based interview, it was feared, would tend to gloss over differences among household members and produce inflated levels of within-household consistency. At the same time, others conjectured — seemingly, just as reasonably — that any differences in the tendency to produce within-household consistency might just as well be attributed to flaws in the person-based design which would lead to false differences among household members where there should be consistency. For example, fluctuations in a respondent's attention level (or variations in other cognitive processes) at different

points in the interview might lead to different interpretations of the same question at different times, thus producing falsely inconsistent answers.

In this section we report on the results of our analysis of within-household consistency differences by type of instrument[10]. Of the six items included in the analysis, three (race, Hispanic origin, and current school enrollment) show no difference in the rate of within-household consistency. The three tables that follow show the significant effects — which, it is important to note, are of inconsistent direction.

i. Citizenship Contrary to the primary concern of ACS subject matter analysts, the person-based format elicited significantly more within-household reporting consistency regarding citizenship than did the topic-based format, as shown in Table 36.

ii. Mobility Table 37 summarizes the within-household consistency results for residency in the current living quarters five years ago. The topic-based instrument identified households in which all members shared the same value on this item significantly more frequently than the person-based instrument.

iii. Non-English "at home" language Finally, as shown in Table 38, compared to person-based interviews, topic-based interviews elicited more frequent within-household consistency on use of a language other than English at home.

Thus, although the analysis identified some significant results, we find no evidence for a consistent effect of instrument type on uniform reporting of various characteristics among all household members. Of course, without validating data of some sort, even had we found more within-household consistency with one instrument or the other, we would still be unable to draw any solid conclusions about the implications of this difference for data quality differences.

4. Additional Research

The initial review of the results of the person/topic experiment suggests mostly quite positive outcomes for the topic-based design. Some lingering questions remain, however, which future research will need to address.

One question concerns the meaning of the observed differences in some of the response distributions, and the differences in the tendency of the two designs to elicit consistent responses from all household members on certain characteristics. The major issue here, of course, is whether these response differences imply data quality differences, and, if so, which instrument produces higher quality data. We will carry out additional analyses on the existing data to try to better understand these phenomena, and will also try to design data quality assessments into future experiments

[10]Again, we thank Gregg Diffendal (CMO) for producing the within-household consistency results and for carrying out the analysis of them

comparing the two designs, in order to provide more definitive evidence on the existence and direction of any data quality differences.

A second high-priority question concerns one aspect of the item nonresponse results in the field experiment. Although the vast majority of item nonresponse differences favored the topic-based design, two instances in which the topic-based instrument produced significantly more item nonresponse were on questions about wage/salary income receipt and amounts, and total income. Because of the importance of income data to the ACS, these particular results are of some concern to CM staff, even though they run counter to the overall nonresponse results. We need to carry out additional research with the existing data to try to assess the extent to which these differences might be due to sample differences, instrument design flaws, or some other factors extrinsic to the topic-based design; or whether they are, in fact, an inherent weakness of a topic-based interview. If the elevated nonresponse to income items proves robust in additional tests of the topic-based design, we will then need to develop new strategies to counteract the effect while maintaining the other advantages of the topic-based approach.

Finally, as noted earlier, the CATI operation is only the first stage of the followup of mail nonresponse in the ACS. CATI in fact produces only about half of all of the ACS interviews obtained from mail nonrespondent households; the remainder are obtained via personal-visit CAPI interviewing. Therefore, it is important to determine whether the beneficial effects of the topic-based approach observed in the CATI setting carry over to a CAPI mode of administration. Pending funding approval, we expect that this will be a major focus of the next stage in the testing and development of a topic-based approach to the ACS mail nonresponse followup interviewing.

5. Summary and Conclusions

The Census Bureau conducts a number of household demographic surveys which gather information about all members of sampled households from a single household respondent. These surveys are typically "person-based" in their design — that is, they ask the complete set of interview questions for each household member in turn, recycling through the interview sequence as many times as there are eligible members of the household. In recent years, the automation of survey instruments has opened possibilities for new questionnaire design options, including a "topic-based" approach, which completes each interview question (or topic) for all household members before proceeding to the next question.

In a large-scale questionnaire design experiment, conducted during the 1997 test of the Census Bureau's American Community Survey (ACS) CATI followup interview, we tested a person-based ACS CATI instrument against a topic-based instrument over a two-month interviewing period. The results of the experiment suggest many advantages of the topic-based design. Compared to the more traditional person-based approach, the topic-based ACS interview achieved a higher response rate, a lower refusal rate, more favorable evaluations from interviewers and respondents, generally lower rates of item nonresponse, and reduced interview length. The positive outcomes derived from the topic-based approach offer strong justification for a change to the topic-based design in the ACS

mail nonresponse followup system, a position we have argued both informally and in a summary report to CM staff (Moore, 1998).

Acknowledgments

We gratefully acknowledge the important contributions of the many people who have assisted with the research described in this report, especially including the following:

- Wendy Davis, Elaine Fansler, Meredith Lee, Lorraine Randall, and Dawn Von Thurn, for their help in designing and carrying out the pilot study which preceded the CATI field experiment;

- Bonnie Carver, Annetta DePompa, Gregg Diffendal, Linda Hiner, Dale Lewis, Glenn Schneider, Richard Smiley, and Claudia West, for their help in designing, implementing and analyzing the results of the CATI field experiment;

- Tom Melaney and Claudia West at Census headquarters, and Kimberly Clark, Kathy Sayler, and their staff at the Hagerstown Telephone Center, for their work on the behavior coding phase of the study;

- the tireless and dedicated ACS/CATI supervisors and interviewing staff at the Jeffersonville Telephone Center; and

- Annetta DePompa and Gregg Diffendal, for their helpful comments on early drafts of this report.

References

Fowler, F. and Cannell, C. (1996). Using behavior coding to identify cognitive problems with survey questions. In N. Schwarz and S. Sudman (eds.), Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass.

Groves, R. and Couper, M. (in press). Nonresponse in Household Interview Surveys. New York: John Wiley & Sons.

McGuire, W. (1968). The nature of attitudes and attitude change. In G. Lindzey and E. Aronson (eds.), The Handbook of Social Psychology, 2nd Edition, Volume 3. Reading, MA: Addison-Wesley.

Moore, J. (1996). Person- vs. topic-based design for computer-assisted household survey instruments. Paper presented at the International Conference on Computer-Assisted Survey Information Collection, San Antonio, TX, December 11-14, 1996.

Moore, J. (1998). ACS/CATI person-based vs. topic-based field experiment — Evaluation summary and recommendation. U.S. Bureau of the Census memorandum to Larry McGinn, February 19, 1998.

Attachment 1

Items for which analysis of the response distributions showed no significant difference between the two instrument types (see section 3.5.2):

1. Basic (100%) Demographic Items

relationship to reference person (B2)
sex (B3)
age (B4, collapsed into 5-year categories)
marital status (B5)
Hispanic origin (B6a)

2. Detailed Demographic Items

highest grade/degree (P11, collapsed into 9 categories: no school/pre- school, K through 6th grade, 7th through 12th grade (no diploma), high school graduate or equivalent, some college (no degree), Vo/Tech/Bus school degree, associate degree, bachelors degree, masters/professional/PhD)
live here 5 years ago (P13a)
hear normal conversation (P15b)
walk 1/4 mile (P15c)
go outside alone (P16a)
long-lasting physical or mental condition (P16b)
number of babies (P17, collapsed into 7 categories: 0, 1... 5, 6+)
ever served in armed forces (P18a)
hours worked last week (P22, collapsed into 3 categories: <35 hours, 35-40 hours, 41+ hours)
type of transportation to work (P24)
on layoff last week (P28a1)
looking for work last 4 weeks (P29)
available to start a job if offered (P30b)
when did [name] last work (P31)
weeks worked last 12 months (P32_1, collapsed into 4 categories: 12 or fewer weeks, 13-26 weeks, 27-36 weeks, 37-52 weeks)
usual hours worked per week (P33, collapsed into 3 categories: 34 or fewer hours, 35-40 hours, 41+ hours)
type of business — manufacturing, wholesale, etc. (P37)
receive any net rental/royalty/estate-trust income (P40c3)
receive any social security/railroad retirement (P40d1)
receive any retirement/survivor/disability pensions (P40e1)
receive any SSI/AFDC/public assistance/welfare (P40f1)
receive any other (VA, UI, child support, alimony) (P40g1)

Technical Appendices
(available upon request from the Center for Survey Methods Research)

Technical Appendix 1:	Pilot test respondent debriefing questionnaire
Technical Appendix 2:	Pilot test respondent debriefing interview protocol
Technical Appendix 3:	Pilot test coding sheets for respondents' verbal and non-verbal displays
Technical Appendix 4:	Interviewer debriefing moderator's guide
Technical Appendix 5:	Behavior coding implementation details and coding scheme

Tables

Table 1: Pilot Test Interview Length by Instrument Type	Average Length of Interview (minutes)	
	Per Interview	Per Adult Interviewed
Person-Based	32.7	11.4
T1 Topic-Based	40.2	12.5
T2 Topic-Based	27.2	10.3

Table 2: Pilot Test Debriefing Results by Instrument Type	Average Debriefing Score		
	Repetitive? (1=not repetitive; 5=very repetitive)	Impatient? (1=not impatient; 5=very impatient)	Prefer the "other" type of interview structure? (% "yes")
Person-Based	4.2	2.3	79%
T1 Topic-Based	3.7	2.2	33%
T2 Topic-Based	2.8	1.6	27%

Table 3: Pilot Test Behavior Observation Results by Instrument Type	Average Number of Observed "Displeasure" Displays per Interview		
	Boredom/ Fatigue	Annoyance/ Desire to Speed-Up the Interview	Confusion
Person-Based	15.9	3.2	3.0
T1 Topic-Based	19.3	3.9	3.1
T2 Topic-Based	0.4	2.0	2.3

Table 4: Workload for the Person/Topic Field Test by Interview Month and Instrument Type		October		November		TOTAL	
		Person-Based	Topic-Based	Person-Based	Topic-Based	Person-Based	Topic-Based
Total Cases Sent to CATI Followup		1607	1607	1338	1337	2945	2944
Ineligible for CATI*:	Late Mail Returns	233	265	146	156	379	421
	Other Ineligible	299	315	298	302	597	617
	Total	532	580	444	458	976	1038
Eligible for CATI Followup Interview		1075	1027	894	879	1969	1906

*Note: Late Mail Return cases were generally removed from the CATI workload before a call was completed. "Other" ineligible cases consist mostly of attempted cases for which the listed telephone number failed to reach the correct address, as well as a small number of cases which were found to be businesses or otherwise non-residential addresses.

Table 5: Household Response Outcomes by Interview Month and Instrument Type	October		November		TOTAL	
	Person-Based	Topic-Based	Person-Based	Topic-Based	Person-Based	Topic-Based
Total Eligible CATI Cases	100% (1075)	100% (1027)	100% (894)	100% (879)	100% (1969)	100% (1906)
Completed Interviews	55.4% (596)	58.4% (600)	57.7% (516)	63.0% (554)	56.5% (1112)	60.5% (1154)
Refusals	15.3% (164)	11.2% (115)	16.8% (150)	15.1% (133)	15.9% (314)	13.0% (248)
Other Noninterviews	29.3% (315)	30.4% (312)	25.5% (228)	21.8% (192)	27.6% (543)	26.4% (504)

Table 6: Average Interview Length in Minutes by Instrument Type and Household Size	Household Size					Mean Interview Length
	2 Persons	3 Persons	4 Persons	5 Persons	6+ Persons	
Person-Based	29.3 (320)	34.7 (202)	36.5 (171)	39.9 (87)	47.2 (75)	34.7 (855)
Topic-Based	27.6 (350)	32.5 (197)	34.7 (190)	36.6 (90)	51.0 (54)	32.6 (881)
	t=5.43 p<.001	t=5.27 p<.001	t=3.56 p<.001	t=4.26 p<.001	t=1.39 n.s.	t=3.99 p<.001

Table 7: Interviewers' "Like" and "Dislike" Reports by Interview Month and Instrument Type				
Part A: Interviewers' "Like" Responses	What did you like about the [person/topic]-based instrument?			
	October		November	
	% of Interviewers who Provided "Like" Comments	Average Number of "Likes" per Commenter	% of Interviewers who Provided "Like" Comments	Average Number of "Likes" per Commenter
Person-based	55% (22)	1.0 (12)	50% (12)	1.2 (6)
Topic-based	93% (14)	1.4 (13)	100% (20)	1.8 (20)
Part B: Interviewers' "Dislike" Responses	What did you dislike about the [person/topic]-based instrument?			
	October		November	
	% of Interviewers who Provided "Dislike" Comments	Average Number of "Dislikes" per Commenter	% of Interviewers who Provided "Dislike" Comments	Average Number of "Dislikes" per Commenter
Person-based	86% (22)	1.3 (19)	83% (12)	1.7 (10)
Topic-based	43% (14)	1.0 (6)	50% (20)	1.1 (10)

Table 8: Interviewers' Assessments of Instrument Design Advantages by Interview Month and Instrument Type	Please rate the [person/topic]-based instrument. Did organizing the questions so that all questions were asked [for one person before moving on to the next person / for all persons for a given topic before moving on to the next topic] ...			
	% Strongly agree or agree			
	October		November	
	Person-Based (n=22)	Topic-Based (n=14)	Person-Based (n=12)	Topic-Based (n=19)
... make interviews flow quickly	36%	92%*	8%	95%
... make interviews flow naturally	27%	100%	17%	90%
... help you become familiar with the organization of the instrument	64%	85%*	8%	84%
... make it easy to conduct an interview	45%	100%	42%	90%
... make it easier to probe for more correct answers	64%	86%	25%	65%

*Note: One October topic-based debriefing questionnaire lacked a response to the starred items, so the denominator for calculating the cell percentages is 13 cases, rather than 14.

Table 9: Interviewers' Assessments of Instrument Performance in Various Kinds of House-holds by Instrument Type	Please rate how you felt the [person/topic]-based instrument performed with different types of households...			
	% Reporting "Very Well" or "Well"			
	October		November	
	Person-Based (n=24)	Topic-Based (n=14)	Person-Based (n=12)	Topic-Based (n=20)
Single Person Households	100% (24)	85% (13)	100% (11)	65% (20)
2-3 Person Households	82% (22)	100% (14)	50% (12)	65% (20)
4+ Person Households	9% (22)	93% (14)	9% (11)	80% (20)
Households with Children	49% (21)	100% (14)	36% (11)	95% (20)
Households with Unrelated Persons	28% (18)	69% (13)	40% (10)	50% (20)
Reluctant/Unenthusiastic Respondents	0% (22)	57% (9)	9% (11)	65% (20)
Elderly Respondents	32% (22)	79% (14)	9% (11)	65% (20)

Table 10: Interviewers' Expressed Interview Design Preferences by Interview Month and Instrument Type	(October) During the past month, you have conducted the interviews by [asking all the questions for one person / taking each topic and asking the question for everyone] before going on to the next [person / topic]. If instead, you had [taken a topic and asked the question of everyone / asked all of the questions for one person] before going on to the next [questions / person], do you think that would make the interview go more smoothly? (November) During the past few months, you have conducted the interviews using both the topic-based and person-based instrument. Which instrument do you prefer and why?			
	October		November	
	Person-Based (n=21)	Topic-Based (n=14)	Person-Based (n=18)	Topic-Based (n=12)
[Preferred "Own" Instrument]	24%	86%	17%	100%
[Preferred "Other" Instrument]	76%	14%	83%	0%

Table 11: Exclusion of Cases from the Respondent Debriefing Analysis (see text for details)	Instrument Type		
	Person-Based	Topic-Based	Total
TOTAL, COMPLETED INTERVIEWS	1112	1154	2266
- No permanent residents (temporarily occupied units)	14	14	28
= Sub-Total, Occupied Housing Units	1098	1140	2238
- One-person households	243	259	502
= Sub-Total, 2+ Person Households	855	881	1736
- Non-continuous interviews (no debriefing)	141	95	236
= FINAL ANALYSIS SAMPLE	714	786	1500

Table 12: Perceived Interview Length by Instrument Type	How did you feel about the length of the survey? Would you say it was too short, about the right length, or too long?				
	% Too Short 1	% About Right 2	% Too Long 3	Mean (3-pt. scale)	(missing)
Person-Based	0.4% (3)	55.4% (391)	44.2% (312)	2.44	— (8)
Topic-Based	1.0% (8)	57.7% (448)	41.3% (321)	2.40	— (9)
	χ^2=2.88, n.s.			t=1.32, n.s.	

Table 13a: Felt Boredom/ Impatience by Instrument Type	Did you stay interested all the way through the interview, or did you get bored or impatient sometimes?		
	% Stayed Interested	% Got Bored or Impatient	(missing)
Person-Based	57.7% (408)	42.3% (299)	— (7)
Topic-Based	64.3% (500)	35.7% (278)	— (8)
	t=2.59, p<.01		

Table 13b: Level of Felt Boredom/ Impatience by Instrument Type	Did you stay interested all the way through the interview, or did you get bored or impatient sometimes? + *(if "bored/impatient")* Would you say you got a little bored or impatient, or very bored or impatient?				
	% Stayed Interested 1	% Got A LITTLE Bored/Impatient 2	% Got VERY Bored/Impatient 3	Mean (3-pt. scale)	(missing)
Person-Based	57.7% (408)	33.8% (239)	8.5% (60)	1.51	— (7)
Topic-Based	64.3% (500)	25.2% (196)	10.5% (82)	1.46	— (9)
	χ^2=13.62, 2 df, p<.001			t=1.31, n.s.	

Note: The 3 respondents (1 person-based; 2 topic-based) who reported being "bored," but who failed to provide a useable response to the followup item, are included in the "a little bored" category.

Table 14: Perceived "Smoothness" of the Interview by Instrument Type	In general, do you feel that the interview moved along smoothly, or did it sometimes seem to jump around?		
	% Moved Smoothly	% Jumped Around	(missing)
Person-Based	97.6% (691)	2.4% (17)	— (6)
Topic-Based	95.1% (742)	4.9% (38)	— (6)
	t=2.57, p<.05		

Table 15a: Perceived "Repetitiousness" of the Interview by Instrument Type	Did the interview questions seem repetitious to you?		
	% Yes (repetitious)	% No (NOT repetitious)	(missing)
Person-Based	58.6% (413)	41.4% (292)	— (9)
Topic-Based	55.1% (430)	44.9% (350)	— (6)
	t=1.34, n.s.		

Table 15b: Perceived "Repetitiousness" of the Interview by Instrument Type and Presence of Non-Relatives	Did the interview questions seem repetitious to you?		
	% Yes (repetitious)	% No (NOT repetitious)	(missing)
Households with ALL RELATED PERSONS:			
Person-Based	59.3% (388)	40.7% (266)	— (8)
Topic-Based	54.4% (396)	45.6% (332)	— (5)
Households with ONE OR MORE NON-RELATIVES:			
Person-Based	49.0% (25)	51.0% (26)	— (1)
Topic-Based	65.4% (34)	34.6% (18)	— (1)
	test for significant interaction: t=2.11, p<.05		

Table 15c: Level of Perceived "Repetitiousness" by Instrument Type and Presence of Non-Relatives	Did the interview questions seem repetitious to you? + *(if "yes (repetitious)")* Would you say they were somewhat repetitious, or very repetitious?				
	% Not Repetitious 1	% Somewhat Repetitious 2	% Very Repetitious 3	Mean (3-pt. scale)	(missing)
Households with ALL RELATED PERSONS:					
Person-Based	40.7% (266)	36.9% (241)	22.5% (147)	1.82	— (8)
Topic-Based	45.6% (332)	38.1% (277)	16.4% (119)	1.71	— (5)
Households with ONE OR MORE NON-RELATIVES:					
Person-Based	51.0% (26)	35.3% (18)	13.7% (7)	1.63	— (1)
Topic-Based	34.6% (18)	50.0% (26)	15.4% (8)	1.81	— (1)
Note: The 4 respondents (2 person-based; 2 topic-based) who reported finding the questions "repetitious," but who failed to provide a useable response to the followup item, are included in the "somewhat repetitious" category.				test for signif. interaction: t=1.90, p<.10	

Table 16a: Expressed Preference for the "Other" Interview Approach by Instrument Type	Let me ask you about the ordering of the interview questions. There are two ways I could have conducted the interview. With you I [took each topic — date of birth, for example — and asked the questions for everyone / asked all of the questions for one person] before going to the next [topic / person]. If instead I had [finished all the questions for one person / taken each topic — date of birth, for example — and finished the questions for everyone] before going to the next [person / topic], do you think that would have made the interview go more smoothly?		
	% Yes (other design would have gone more smoothly)	% No (other design would NOT have gone more smoothly)	(missing)
Person-Based	34.7% (236)	65.4% (445)	— (33)
Topic-Based	10.2% (78)	89.8% (690)	— (18)
	t=11.52, p<.0001		

Table 16b: Expressed Preference for the "Other" Interview Approach by Instrument Type and Presence of Non-Relatives	... would [the other design] have made the interview go more smoothly?		
	% Yes	% No	(missing)
Households with ALL RELATED PERSONS:			
Person-Based	35.5% (224)	64.5% (407)	— (31)
Topic-Based	9.8% (70)	90.2% (646)	— (17)
Households with ONE OR MORE NON-RELATIVES:			
Person-Based	24.0% (12)	76.0% (38)	— (2)
Topic-Based	15.4% (8)	84.6% (44)	— (1)
	test for significant interaction: t=2.12, p<.05		

Table 17a: Perceived Responsibility to Answer "Surveys like this" by Instrument Type	Do you agree or disagree with this statement: "Everyone has a responsibility to answer surveys like this."		
	% Agree	% Disagree	(missing)
Person-Based	74.4% (512)	25.6% (176)	— (26)
Topic-Based	66.1% (495)	33.9% (254)	— (37)
	t=3.47, p<.001		

Table 17b: Strength of Perceived Responsibility to Answer "Surveys Like This"	Do you agree or disagree with this statement: "Everyone has a responsibility to answer surveys like this." + Would you say that you strongly (agree/disagree), or do you just (agree/disagree)?					
	% Strongly Agree 1	% Agree 2	% Disagree 3	% Strongly Disagree 4	Mean (4-pt. scale)	(missing)
Person-Based	29.2% (201)	45.2% (311)	14.1% (97)	11.5% (79)	2.08	— (26)
Topic-Based	24.7% (185)	41.4% (310)	20.3% (152)	13.6% (102)	2.23	— (37)
	χ^2=13.17, 3 df, p<.005				t=2.97, p<.005	

Note: The 23 respondents (13 person-based; 10 topic-based) who responded to the initial "responsibility" item, but who failed to provide a valid response to its followup, are included in either the "agree" or "disagree" category, depending on their initial response.

Table 18a: Perception that "Surveys like this" are a Waste of Time	And finally — Do you agree or disagree with this statement: "Surveys like this one are a waste of people's time."		
	% Agree	% Disagree	(missing)
Person-Based	18.8% (126)	81.3% (546)	— (42)
Topic-Based	24.1% (177)	76.0% (559)	— (50)
	t=2.43, p<.05		

Table 18b: Strength of Perception that "Surveys Like This" are a Waste of Time by Instrument Type	And finally — Do you agree or disagree with this statement: "Surveys like this one are a waste of people's time." + Would you say that you strongly (agree/disagree), or do you just (agree/disagree)?					
	% Strongly Agree 1	% Agree 2	% Disagree 3	% Strongly Disagree 4	Mean (4-pt. scale)	(missing)
Person-Based	8.9% (60)	9.8% (66)	53.9% (362)	27.4% (184)	3.00	— (42)
Topic-Based	11.7% (86)	12.4% (91)	49.6% (365)	26.4% (194)	2.91	— (50)
	χ^2=5.99, 3 df, n.s.				t=1.92, p<.10	

Note: The 31 respondents (14 person-based; 17 topic-based) who responded to the initial "waste of time" item, but who failed to provide a valid response to its followup, are included in either the "agree" or "disagree" category, depending on their initial response.

Table 19: Cases Included in the Behavior Coding Analysis by Interview Month and Instrument Type	Interview Month		Total, both months
	October	November	
Person-Based	54	40	94
Topic-Based	45	43	88
Total, both instruments	99	83	182

Table 20: Behavior Coding Results by Instrument Type	Person-Based	Topic-Based
Number of Interviews Coded	94	88
Number of "On-Path" Questions: Minimum / Maximum	38 / 367	50 / 352
Average	143.2	141.5
PART 1: INTERVIEWER BEHAVIORS		
Ave. % Exact Wording + Slight Wording Change	72.5	70.7
" Correct Verification	5.5	6.0
" Multiple Verification	0.1	0.4
" Major Wording Change	13.8	12.6
" Multiple Question Asking	0.3	0.6
" Silent Verification	2.7	3.9
" Omission	5.0	5.7
PART 2: RESPONDENT BEHAVIORS		
Ave. % Adequate Answer (full question)	76.0	75.6
" Adequate Answer (break-in)	3.0	2.4
" Inadequate Answer	9.4	11.5
" Break-In (excluding adequate answer break-ins)	2.4	0.6
" Qualified Answer	3.4	3.3
" Request for Clarification	3.0	3.9
" Don't Know	2.1	1.9
" Refusal	0.1	0.4
" (?? — garbled response)	0.5	0.4

Table 21: Behavior Coding Match Between Interviewer's Questionnaire Entry and Respondent's Report by Instrument Type	Did the interviewer's entry match the respondent's report?		
	Ave. % Yes (Entry/Report Match)	Ave. % No (Entry/Report MISmatch)	Ave. % Don't Know
Person-Based	92.9	3.4	3.8
Topic-Based	91.5	4.2	4.3
	t=0.35, n.s.		

Table 22: Selected Behavior Coding Results by Household Size and Instrument Type	Number of People in the Household					
	2		3 - 4		5+	
	Person-Based (n=37)	Topic-Based (n=36)	Person-Based (n=41)	Topic-Based (n=30)	Person-Based (n=16)	Topic-Based (n=22)
PART 1: INTERVIEWER BEHAVIORS						
Ave. % Exact/Slight Change	77.2	74.8	70.3	73.0	67.4	61.1
" Major Change	12.4	12.9	14.8	9.8	14.3	15.8
PART 2: RESPONDENT BEHAVIORS						
Ave. % Adequate Answer	78.0	77.1	79.0	80.6	81.2	76.0
" Inadequate Answer	9.5	11.5	9.9	10.6	8.0	12.7

Table 23: Overall Response Completeness by Instrument Type	Ave. % of Non-Missing Items Per Interviewed Person
Person-Based	95.9 (2866)
Topic-Based	96.0 (2866)
	t=0.47, n.s.

Table 24:
Item Nonresponse Rates for "High Nonresponse" Items* by Instrument Type
(* "High nonresponse" = at least one instrument's nonresponse rate exceeds 2%)

Item	Nonresponse Rate ("on-path" n)		Chi-square
	Person-Based	Topic-Based	
(FQ7) Is [name's] race White, Black, Eskimo, Aleut, American Indian, Asian or Pacific Islander, some other race, or multiracial?	2.0% (2866)	1.4% (2866)	3.09 p<.10
(FQ9) When did [name] come to live in the United States?	17.6% (302)	9.1% (409)	11.36 p<.001
(FQ11) What is the highest degree or grade of school [name] completed?	3.2% (2723)	2.0% (2728)	8.42 p<.01
(FQ12) What is [name]'s ancestry?	9.8% (2866)	7.4% (2866)	11.19 p<.001
(FQ13b) Where did [name] live 5 years ago?	6.2% (979)	2.6% (929)	14.90 p<.001
(FQ14b) What is this language [other than English spoken at home]?	12.9% (380)	2.7% (438)	30.41 p<.001
(FQ14c) How well does [name] speak English - very well, well, not well, not at all?	13.2% (380)	3.0% (438)	29.72 p<.001
(FQ15a) Does [name] have any difficulty seeing words and letters in ordinary newspaper print ... ?	2.0% (2638)	1.0% (2629)	9.52 p<.005
(FQ15b) Does [name] have any difficulty hearing what is said in a normal conversation ... ?	2.0% (2638)	1.2% (2629)	5.78 p<.05
(FQ15c) Does [name] have any difficulty walking a quarter of a mile — three city blocks?	2.2% (2638)	1.3% (2629)	5.28 p<.05
(FQ16a) Does [name] have a long-lasting physical or mental condition that makes it difficult for him/her to go outside the home alone ... ?	2.1% (2110)	0.9% (2164)	10.72 p<.001
(FQ16b) Does [name] have a long-lasting physical or mental condition that prevents him/her from working at a job or business?	2.2% (2110)	0.9% (2164)	12.80 p<.001
(FQ17) How many babies has [name] ever had?	2.9% (1066)	1.1% (1097)	9.13 p<.005
(FQ18) Has [name] ever served on active duty in the U.S. Armed Forces, the military reserves, or the National Guard?	2.1% (2110)	0.7% (2164)	14.74 p<.001
(FQ19) When did [name] serve on active-duty in the U.S. Armed Forces?	17.3% (312)	7.3% (274)	13.24 p<.001

Table 24: Item Nonresponse Rates (continued)	Person-Based	Topic-Based	χ^2
(FQ20) In total, how much time has [name] spent on active duty in the Armed Forces? (a) How many years? (b) How many additional months?	20.8% (312)	11.0% (274)	10.49 p<.001
Note: In the CATI instruments this item is split into two parts, as shown. The item as a whole was coded as missing only if both parts (a) and (b) were missing.			
(FQ21) Last week did [name] do any work for pay or profit?	2.5% (2110)	1.0% (2164)	14.20 p<.001
(FQ22) Last week, how many hours did [name] actually work at all jobs?	2.9% (1326)	2.7% (1433)	0.21 n.s.
(FQ23) Last week, where did [name] work?	13.2% (1326)	14.7% (1433)	1.33 n.s.
(FQ26) Last week, what time did [name] usually leave home to go to work?	6.2% (1297)	9.1% (1403)	7.92 p<.005
(FQ27) Last week, how many minutes did it usually take [name] to get from home to work?	5.2% (1297)	6.8% (1403)	3.34 p<.10
(FQ28a) Last week, was [name] on layoff from a job?	6.6% (784)	3.6% (731)	7.33 p<.01
(FQ28b) Is this layoff temporary or permanent?	3.3% (30)	12.9% (31)	1.86 n.s.
(FQ28c) Last week, was [name] temporarily absent from a job or business?	6.9% (754)	3.6% (700)	8.00 p<.005
(FQ29) Has [name] been looking for work during the last 4 weeks?	7.0% (727)	3.8% (681)	6.95 p<.01
(FQ31) When did [name] last work, even for a few days?	7.5% (784)	5.3% (731)	3.00 p<.10
(FQ32) During the past 12 months, in how many weeks did [name] work, even for a few hours?	6.9% (1562)	5.2% (1607)	3.96 p<.05
(FQ33) How many hours did [name] usually work per week?	7.1% (1562)	4.8% (1607)	7.61 p<.01
(FQ34) I am going to read 5 categories. Please pick the one that best describes who [name] works/worked for ... [private organization/company, government, US Armed Forces, self-employed, working without pay in a family business]	4.2% (1562)	2.6% (1607)	6.80 p<.01
(FQ35) What is/was the name of [name]'s company or employer?	7.9% (1562)	9.7% (1607)	3.10 p<.10

Table 24: Item Nonresponse Rates (continued)	Person-Based	Topic-Based	χ^2
(FQ36) What kind of business or industry is/was this?	4.7% (1562)	4.9% (1607)	0.06 n.s.
(FQ37) Is this business mainly ... [manufacturing, wholesale, retail, other]?	4.3% (1562)	4.6% (1607)	0.19 n.s.
(FQ38) What kind of work was [name] doing at this job?	4.4% (1562)	5.1% (1607)	0.82 n.s.
(FQ39) What were [name]'s most important activities or duties?	6.6% (1562)	6.7% (1607)	0.02 n.s.
(FQ40a1) Did [name] receive any wages or salary income from an employer? (FQ40a2) Did [name] receive any income such as commissions, bonuses, or tips from an employer?	22.0% (1704)	25.0% (1762)	4.39 p<.05
Note: In the CATI instruments receipt of any wage/salary income (40a1) was split off from receipt of any commissions/bonuses/tips (40a2), as shown above; furthermore, each was followed by its own "amount" followup (as appropriate). The item as a whole was coded as missing if any part was missing.			
(FQ40b) Did [name] receive any self-employment income?	7.2% (1704)	7.8% (1762)	0.47 n.s.
Note: This "receive any self-employment" item was followed by an "amount" item if the response to the "receive any" item was "yes." The item as a whole was considered missing if either part was missing. The same design and rules apply to items 40c1 through 40g, below.			
(FQ40c1) Did [name] receive any income from interest or dividends?	15.9% (2110)	17.1% (2164)	1.24 n.s.
(FQ40c2) Did [name] receive any net rental income, royalty income, or income from estates and trusts?	6.1% (2110)	6.2% (2164)	0.01 n.s.
(FQ40d) Did [name] receive any Social Security or Railroad Retirement benefits?	7.82% (2110)	8.6% (2164)	0.76 n.s.
(FQ40e) Did [name] receive any retirement, survivor, or disability pensions?	6.5% (2110)	6.1% (2164)	0. 28 n.s.
(FQ40f) Did [name] receive any Supplemental Security Income (SSI), Aid to Families with Dependent Children (AFDC), or other public assistance or public welfare payments?	4.7% (2110)	5.2% (2164)	0.43 n.s.
(FQ40g) Did [name] receive income on a regular basis from any other source such as the Veteran's Administration (VA) payments, unemployment compensation, child support or alimony, etc.?	5.4% (2110)	5.3% (2164)	0.00 n.s.
(FQ41) What was [name]'s total income during the past 12 months?	22.5% (2110)	28.9% (2164)	23.03 p<.001

Table 24: Item Nonresponse Rates (continued)	Person-Based	Topic-Based	χ^2
Note: An incomplete CATI implementation may have worked to the disadvantage of the topic-based instrument in the case of item 41. The intent was to sum all reported amounts and present this total to respondents for verification; however, this function was inoperable in the instrument versions used in this test. Person-based interviewers could (and occasionally did, according to observers) keep a written tally as they proceeded through the questions, and offer their own version of a verification task at item 41 if the respondent hesitated to provide a total amount unaided. Because of the interleaving of the income reports of all adult household members, the topic-based structure rendered it almost impossible for interviewers to compensate for the absent automated tally in this way.			

Table 25: Reported Race by Instrument Type	Is [name's] race White, Black, Eskimo, Aleut, American Indian, Asian or Pacific Islander, some other race, or multiracial?					
	White	Black	Esk./Aleut/ Amer. Ind.	Asian/ Pac. Is.	Other	Multi-Racial
Person-Based	2190 (.71)	565 (.18)	18 (.01)	68 (.02)	156 (.05)	77 (.03)
Topic-Based	2180 (.71)	540 (.18)	10 (*)	137 (.04)	136 (.04)	82 (.03)
	chi-square = 27.6, 5 df, $p < .001$					

Note that the version of the race question tested in the 1997 experiment has since been superceded with slightly different response categories and a format which permits multiple entries.

Table 26: Reported Citizenship by Instrument Type	Is [name] a citizen of the United States?				
	Yes, born in the U.S.	Yes, born in PR, Guam, etc.	Yes, born abroad of Am. parents	Yes, by naturalization	No, not a citizen
Person-Based	2789 (.91)	16 (.01)	16 (.01)	135 (.04)	104 (.03)
Topic-Based	2694 (.87)	32 (.01)	24 (.01)	175 (.06)	177 (.06)
	chi-square = 32.4, 4 df, $p < .001$				

Table 27: Reported Recent School Enrollment by Instrument Type	At any time in the past 3 months, was [name] attending a school or college?	
	Yes	No
Person-Based	927 (.32)	1988 (.68)
Topic-Based	857 (.29)	2108 (.71)
	chi-square = 5.8, 1 df, $p < .05$	

Table 28: Reported Use of a Language Other than English at Home by Instrument Type	Does [name] speak a language other than English at home?	
	Yes	No
Person-Based	350 (.12)	2482 (.88)
Topic-Based	451 (.16)	2424 (.84)
	chi-square = 13.1, 1 df, $p < .001$	

Table 29: Reported Reading Disability by Instrument Type	Does [name] have any difficulty seeing words and letters in ordinary newspaper print ...?	
	Yes	No
Person-Based	127 (.04)	2702 (.96)
Topic-Based	156 (.05)	2706 (.95)
	chi-square = 2.8, 1 df, $p < .10$	

Table 30: Reported Work for Pay/Profit Last Week by Instrument Type	Last week did [name] do any work for pay or profit?	
	Yes	No
Person-Based	1448 (.63)	853 (.37)
Topic-Based	1573 (.66)	825 (.34)
	chi-square = 3.6, 1 df, p < .10	

Table 31: Reported Number of Co-Riders to Work by Instrument Type	Last week, how many people including [name] usually rode to work together?		
	1	2	3 or more
Person-Based	1138 (.91)	83 (.07)	29 (.02)
Topic-Based	1182 (.88)	131 (.10)	26 (.02)
	chi-square = 8.7, 2 df, p < .05		

Table 32: Reported Employer Category by Instrument Type	Please pick the one [category] that best describes who [name] works for ...				
	Private org. or company	Government	U.S. Armed Forces	Self-employed	Working w/o pay in family business
Person-Based	1233 (.76)	227 (.14)	8 (*)	160 (.10)	5 (*)
Topic-Based	1289 (.75)	208 (.12)	5 (*)	216 (.13)	5 (*)
	chi-square = 8.7, 4 df, p < .10 (% missing = 3.2%)				

Table 33: Reported Receipt of any Wage/Salary Income by Instrument Type	Did [name] receive any wage or salary income from an employer during the past 12 months?	
	Yes	No
Person-Based	1224 (.84)	239 (.16)
Topic-Based	1189 (.81)	273 (.19)
	chi-square = 2.8, 1 df, $p < .10$ (% missing = 23.1%)	

Table 34: Reported Receipt of any Self-Employment Income by Instrument Type	Did [name] receive any self-employment income?	
	Yes	No
Person-Based	114 (.07)	1619 (.93)
Topic-Based	163 (.09)	1628 (.91)
	chi-square = 7.7, 1 df, $p < .005$ (% missing = 7.3%)	

Table 35: Reported Receipt of any Interest/Dividend Income by Instrument Type	Did [name] receive any interest or dividends during the past 12 months?	
	Yes	No
Person-Based	318 (.16)	1653 (.84)
Topic-Based	409 (.20)	1592 (.80)
	chi-square = 12.3, 1 df, $p < .001$ (% missing = 16.8%)	

Table 36: Within-Household Consistency on Citizenship by Instrument Type	Is [name] a citizen of the United States?	
	[The SAME characteristic was reported for all persons]	[The characteristic VARIED across persons]
Person-Based	733 (.86)	115 (.14)
Topic-Based	722 (.83)	150 (.17)
	chi-square = 4.4, 1 df, $p < .05$	

Table 37: Within-Household Consistency on 5-Years-Ago Residence by Instrument Type	Did [name] live in this [house/ apartment/...] 5 years ago?	
	[The SAME characteristic was reported for all persons]	[The characteristic VARIED across persons]
Person-Based	729 (.85)	126 (.15)
Topic-Based	795 (.90)	86 (.10)
	chi-square = 10.0, 1 df, $p < .005$	

Table 38: Within-Household Consistency on Language Spoken at Home by Instrument Type	Does [name] speak a language other than English at home?	
	[The SAME characteristic was reported for all persons]	[The characteristic VARIED across persons]
Person-Based	769 (.91)	79 (.09)
Topic-Based	816 (.94)	56 (.06)
	chi-square = 5.0, 1 df, $p < .05$	