RESEARCH REPORT SERIES
*(Survey Methodology #2004-01)*

# Exploiting Computer Automation to Improve the Interview Process and Increase Survey Cooperation

Jeffrey C. Moore

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

*Report Issued:* August 10, 2004

# Exploiting Computer Automation to Improve the Interview Process and Increase Survey Cooperation[1]

Jeffrey C. Moore
Statistical Research Division
U.S. Census Bureau

**Abstract**: Couper (2002) outlines the "challenges and opportunities" of recent and still-emerging technological developments on the conduct of survey research. This paper focuses on one such development – the use of computer-assisted survey instruments in place of paper-and-pencil questionnaires – and it focuses on one particular opportunity which this development presents: the ability to improve efficiency, "flow," and naturalness, and in general make the interview experience a more pleasant one for all participants, while still controlling question wording and sequencing. Moral arguments can be raised in defense of such efforts; the potential for important practical benefits, including improved survey cooperation, lends more mundane but perhaps more potent support. Although the research literature is surprisingly scant, there is some evidence that improved instrument design can reduce nonresponse. A recent effort by the U.S. Census Bureau to redesign the core instrument for the Survey of Income and Program Participation (SIPP) offers additional support. Motivated in large measure by evidence of increasing unit nonresponse and attrition, the primary goal of the SIPP redesign effort was to improve the interview process, and in particular to seek ways to avoid violations of conversational norms (e.g., Grice, 1975). A great many of the SIPP interview process improvements would not have been feasible without the computerization of the survey instrument. This paper briefly summarizes many of the technology-based changes implemented in the SIPP instrument, and briefly describes a set of field experiments used to develop and refine the new procedures and to evaluate their success in achieving SIPP's redesign goals.

**Keywords**: burden, conversational norms, efficiency, flow, nonresponse/attrition, questionnaire design, respondent-friendly

## I. Interviews as Social Interactions

Abundant everyday evidence suggests that people tend to enjoy talking about themselves, and want to be paid attention to. A minor mystery, then, is why they are not more eager to participate in surveys. Why don't people jump at the chance? Why is survey participation perceived as a chore? Why do so many, in fact, actively avoid it? Recent years have seen substantial advances

---

[1]This paper was the basis for a poster presentation at the 2004 meetings of the American Association for Public Opinion Research, Phoenix, AZ, May 13-16, 2004. A substantially condensed version will appear in the Proceedings of the American Statistical Association, Survey Research Methods Section, forthcoming. Contact: Room 3133-4; Center for Survey Methods Research; Washington, DC 20233-9100; phone: (301) 763-4975; fax: (301) 457-4931; e-mail: jeffrey.c.moore@census.gov.

in understanding survey nonresponse and the factors that influence it (see, e.g., Groves and Couper, 1998; Groves, Dillman, Eltinge, and Little, 2002). But little of this work has addressed the substantial emotional content that often seems to accompany people's avoidance of survey participation, and their inability to find joy in the experience when they do participate. The germ of a solution to the mystery noted above may lie in the latter point. As has long been recognized (e.g., Converse and Schuman, 1974), an interviewer-administered survey interview is not a neutral measurement process – it's a social phenomenon, carried out by people in a close interaction. And as an interaction, an interview is often difficult and unsatisfying to the participants – perhaps, some would argue, inevitably so (Suchman and Jordan, 1990; Shober and Conrad, 1997; Lynch, 2002; Maynard and Schaeffer, 2002).

The survey literature provides much rich evidence of the social nature of survey interviews – see, for example, Suchman and Jordan (1990) and Maynard et al. (2002). But well before these recent, molecular-level, examinations of interviewer-respondent interactions, others have used more traditional research data to reach the same conclusion. Cannell, Marquis, and Laurent (1977), for example, undertook a detailed tally of verbal behavior in an interview, and found a strong tendency for interviewers and respondents to strike a balance with regard to the volume of behavior, even though "interviewers themselves did not have a characteristic behavior level for all interviews" [p. 21]. Cannell et al. posit a "reciprocal cue-searching process," in which each participant is responsive to cues from the other as to the appropriate nature of their interaction. They also note that a rather staggering amount of the behavior of interviewers and respondents is "non-programmed." In one study, for example, Cannell et al. found that fully one-third of all the behaviors they enumerated occurred in the supposed void between a respondent's adequate answer and the interviewer's beginning to read the next question.

Respondents may be particularly ill-prepared to treat an interview as simply a neutral conduit for information transmission. Schegloff (2002) notes the tendency of respondents to apply the rules they have learned in acquiring the skills of normal social conversation to the interview setting; Briggs (1986) further notes that in fact they often overtly reject the interviewer's attempts to impose special conversation rules for the interview. Clark (1979) finds a clear spillover from the rules of social conversation in respondents' frequent habit of answering what they perceive to be the question behind the question, as opposed to the overt, actual question presented to them. Even those few respondents who may be aware of the goals of survey interviews, and the peculiar rules that guide how those goals are to be reached, may have great difficulty overcoming the years of over-learning of the norms and practices appropriate to regular social situations.

But, as Cannell et al. (1977) show, interviewers feel the pull of an interview's social interaction qualities, too. Simple observation also suggests this. Why else would interviewers ever complain about having to ask "stupid" questions? One likely reason is that they are as well-trained as respondents in the rules of social interaction, and just as likely to accept the notion that "conversational etiquette is ... a central feature of proper comportment" (Briggs, 1986, p. 57). Another is that they may also know, consciously or otherwise, that if they administer a bad questionnaire, respondents are unlikely to grant them a full pardon as an innocent message-

bearer.[2] Third, of course, is that interviewers have a stake in keeping respondents happy, so their antennae are always tuned to anything likely to raise a respondent distress signal.

The inescapably social/interactional character of an interview further suggests that strictly objective, number-of-questions or length-of-interview -based definitions of the "burden" of an interview, such as that employed by the Office of Management and Budget in approving federal government surveys, are inadequate, and perhaps misleading. (In fact, research has typically failed to detect the suspected link between questionnaire length and nonresponse – see Bogen (1996), for example.) Of course one can "overstay one's welcome," but it may be that a survey's true burden has less to do with the sheer bulk of questions and more to do with the extent to which those questions are vague and ill-defined, or unnecessary, or repetitive, or have already been answered, or are unanswerable, or which seem to come out of the blue, ill-matched to the understood and agreed-to nature of the survey.

This line of thought offers a possible set of principles or guidelines as to what constitutes good questionnaire design, an issue that is surprisingly rarely considered in survey methodology. We have, going back a half-century and more (e.g., Payne, 1951), well-established and widely accepted criteria for constructing good, individual questions. But there is little in the published literature beyond occasional tricks of the trade (e.g., put sensitive questions at the end), and vague generalities, as to how good questions should be strung together to form a good questionnaire. The underlying principle which may be of use is that a good questionnaire is one which, all else equal, supports a smooth and appropriate social interaction. Fortunately, those factors are also reasonably well established – see, for example Grice's (1975) "conversational maxims," which describe the basic rules and assumptions under which participants in normal social discourse operate. These are the maxims of quantity (make your contribution to the ongoing interaction as informative as required – neither more nor less so than necessary); quality (do not say that which you believe to be false, or that for which you lack reasonable evidence); relation (what you say should be relevant to the ongoing interaction); and manner (what you say should be clear). Unfortunately, questionnaire design is rarely very sensitive to the social interaction qualities of an interview, and thus it is not surprising that survey participants often find those interactions unsatisfying, or frustrating, or actively annoying.

## II. Why Be Concerned About the Interactional Quality of Survey Questionnaires?

There are at least three practical reasons to be concerned about questionnaire design that is insufficiently sensitive to an interview's interactional quality. First: threats to standardization. The goal of maximizing the controlled, standardized administration of survey questions is not

[2]A fundamental difference between actors and the observers of those actors is for the observers to tend to attribute to actors an internal, dispositional responsibility for their behavior (Jones and Nisbett, 1971). For example, Manis et al. (1974) find that even those who are explicitly and obviously playing the role of message-bearer, and delivering the words of others, tend to be seen as endorsing or "owning" their spoken behavior.

well served by stressing to interviewers that they must read every question exactly as worded, and then giving them interview scripts that they cannot possibly read exactly as worded if they hope to maintain any semblance of a normal human interaction. This disconnect can only lead interviewers to doubt that the "read every question exactly as worded" mantra is to be taken seriously, and thus to broaden the circumstances under which they feel free to depart from the scripted question text.

The second reason for concern about questionnaires whose designs cause problems for the interviewer/respondent interaction has to do with threats to data quality. Suchman and Jordan (1990) helped open the eyes of the survey research world to the extent to which such "interactional troubles" can bedevil interviewers and respondents as they attempt to navigate the standardized questionnaire interview task. The primary focus of their work, and of virtually all of the work that has spun off from it (e.g., Maynard et al., 2002; Shober and Conrad, 1997), is the inherent shortcomings of a standardized survey questionnaire as a script for a highly complex and unpredictable social interaction, and the negative impact of those shortcomings on data quality. These researchers focus on the trail from interactional blundering, to misunderstanding and miscommunication, to lower quality data. However, a more motivated data quality impact is also possible. To the extent that an interview runs roughshod over interactional niceties, one might expect that respondents – and perhaps interviewers, too – would de-value the importance of the enterprise, and be less engaged in the task of producing good data. Lowered engagement may lead to "satisficing" (e.g., Krosnick et al., 1996) – producing and recording minimally acceptable replies for the purpose of simply getting through the interview as quickly as possible.

The third practical reason – and the primary focus of this paper – is threats to survey cooperation. All else equal, a survey beset with more "interactional troubles" may also suffer more problems obtaining cooperation. The evidence is clear that interviewers and respondents can't prevent the deeply ingrained rules and mores of "normal" conversational interaction from spilling over into the survey setting. Relative to a smooth and satisfying interaction, an interaction marked by violations of those rules and mores is no doubt more distressing and off-putting, and thus less likely to leave the participants with a desire to repeat the experience.

It is important to bear in mind that there are two immediate participants in a survey interaction. In a first contact interview the potential respondent who answers the door/telephone has no idea as to the nature of the questionnaire that is going to be administered, and whether the quality of the upcoming interview interaction is likely to be high or low due to design attention/inattention. But the interviewer knows, and he or she is hardly immune to the unpleasantness of interactional rough spots. It is not a great stretch to imagine some variability in interviewers' behavior as they attempt to secure respondent cooperation, depending on their expectation of the likely quality of the upcoming interaction. We would expect, however, that the effects of questionnaire design would be especially potent in repeat-visit, longitudinal surveys, where both respondents and interviewers can base their actions on direct experience.

## III. Research Evidence that the Quality of an Interview Interaction Matters for Survey Cooperation

A. Survey Introductions

There is evidence in the questionnaire design research literature that good design *does* matter – that a better designed instrument and procedures (or perhaps a more adept interviewer, who is more capable at smoothing the rough edges of a flawed instrument) can lead to reduced survey nonresponse. One strand of evidence can be found in research on survey introductions and other behaviors that occur in the opening seconds of a survey interaction. Research suggests that it is not so much the specific content of the initial interaction that matters as the form and nature of that interaction – *what* is said seems to be less predictive of nonresponse than *how* it is said.

Houtkoop-Steenstra and van den Bergh (2000), for example, note that variations in survey introductions (along such dimensions as long vs. short introductions, introductions which provide more vs. less detail regarding the survey's content and purpose, introductions which stress (or not) the "social significance" of the survey, the strength of confidentiality assurances, etc.) typically show only minimal effects on nonresponse. Houtkoop-Steenstra and van den Bergh suggest that this is because this research has not varied what <u>really</u> matters to respondents, which is the scriptedness of the introduction. They hypothesize that a read-from-a-script introduction will almost always be perceived as such by respondents, which will signal to them that what follows will not be "real" talk, and thus not a satisfying interaction. On the other hand, an unscripted, "conversational" introduction offers more opportunity for interviewers to be flexible, and to adapt their speech offerings (in both style and in content) to the respondent. This should increase the likelihood that the interaction will be positive, and thus that the respondent will comply with the interview request. And, indeed, Houtkoop-Steenstra and van den Bergh find, in an experiment comparing a conversational introduction versus a standard, scripted one, that the former yields significantly more completed interviews, more appointments, and fewer refusals.

Earlier work by Morton-Williams (1991, 1993) also finds a positive impact of improvised introductions on survey cooperation. Interviewers who were required to use a prepared, scripted introduction experienced more refusals than those who (with appropriate training and guidance) were allowed to improvise. Like Houtkoop-Steenstra and van den Bergh, Morton-Williams also concludes that the script interfered with interviewers' ability to be spontaneous and flexible, and to respond appropriately to the unfolding situation.

Research on "tailoring" survey introductions can be seen in much the same light. "Tailoring" seeks to reduce survey nonresponse by training interviewers to attend carefully to respondents' doorstep comments, distill the essential information contained in them concerning any reluctance to cooperate, and to directly respond to the comments with appropriate counter-arguments. Note the clear parallels in Grice's (*op cit*) maxims that conversational offerings in "normal" speech should be both relevant and informative. This suggests that the power of tailoring derives in some measure from their adherence to conversational norms, and, conversely, that damage can

result from not adhering to those norms. Indeed, a large and growing body of evidence finds that tailoring techniques reduce survey refusals (see, e.g., Couper, 1997; Couper and Groves, 2002; Groves, Cialdini, and Couper, 1992; Groves and Couper, 1998; Groves, McGonagle, and O'Brien, 1999; Mayer and O'Brien, 2001; O'Brien, Mayer, Groves, and O'Neill, 2002).

B. Characteristics of 2nd Wave Nonrespondents in Panel Surveys.

Another strand of evidence pointing to the importance of interaction-sensitive survey procedures comes from quasi-experimental examinations of "attritors" in longitudinal surveys – respondents to the initial survey administration who nonrespond in the next survey wave. Kalton et al. (1990) report a significant association between interviewers' ratings of a positive wave 1 interview experience for the respondent (he/she understood the questions, exhibited good cooperation, seemed to enjoy the interview) and successful completion of the second interview approximately 2-1/2 years later. Lepkowski and Couper (2002) take a different analytical approach to the same data Kalton et al. used, and extend the analysis to another, similar dataset. Their refinements in fact yield quite similar results. They again find significant associations of the wave 1 "quality of experience" variables and wave 2 cooperation for both the re-analyzed Kalton et al. data and the new dataset.

C. Experimental Studies of Alternative Questionnaire Designs

The attrition studies cited above suggest a link between the quality of an initial interview interaction and the likelihood of obtaining a subsequent interview, but their non-experimental design leaves them quite uninformative concerning the causal nature of the nonresponse effect. Such studies are useful for developing insights, and for assessing whether observed associations are consistent with causal hypotheses, but the cornerstone of causal inference is experimentation. This section reviews experimental work linking questionnaire design and survey cooperation.

1. Topic-based vs. person-based questionnaire design

Moore and colleagues (e.g., Moore, 1996; Moore and Moyer, 2002) have conducted research on "topic-based" questionnaire design (see also Couper, Fuchs, Hansen, and Sparks, 1997; and Fuchs, 2002). The topic-based design is relevant for survey questionnaires that gather data about all household members from a single household respondent. A topic-based interview completes one topic for all persons before proceeding to the next topic, e.g.:

> 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]

The topic-based design stands in contrast to a conventional "person-based" design, which in essence completes all topics for one person before proceeding to the next person. A key practical benefit of the topic-based interview is the opportunity it affords for severely truncating the text needed for subsequent question administrations after the first, full presentation, e.g.:

> Is [person1] currently married, widowed, divorced, separated, or has he/she never been married? How about [person2]...? And [person3]...? [etc.]

Moore (1996) first demonstrated the potential benefits of the topic-based design in a small experimental pilot test in which a topic-based questionnaire resulted in shorter interviews; was judged by respondents to be less repetitive than the person-based format, and to have elicited reduced feelings of impatience; and was the preferred format among all experimental subjects, by about a 3-to-1 margin. In addition, interview observers found that those exposed to the topic-based instrument displayed significantly less confusion than their person-based counterparts, gave less evidence of annoyance or wanting to speed up the interview, and exhibited fewer displays of boredom or fatigue.

The pilot test was followed by a large-scale field experiment carried out in the context of a pretest of the Census Bureau's American Community Survey[3] CATI nonresponse follow-up procedures – see Moore and Moyer (2002) for details. Test results showed that topic-based interviews were significantly shorter than person-based (see also Colosi, 2001). The topic-based design was clearly preferred by both interviewers and respondents. More important for present purposes, the topic-based instrument also yielded improved survey cooperation – a significantly higher response rate (60.5%, vs. 56.5% in the person-based treatment group), due primarily to a significantly lower refusal rate (13.0%, vs. 15.9% for the person-based group). As Moore and Moyer note, this result, in a one-time survey, underscores the critical yet subtle role played by interviewer behavior in determining survey cooperation outcomes: "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" [p. 12]. But the design of the experiment did fully reveal to interviewers the instrument treatment to be used. Thus, the authors speculate that interviewers may not have "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" [p 12].

2. Household screener vs. person-level questions

A second example of experimentation with the impact of questionnaire design on survey cooperation is reported in Hess et al. (2001). These researchers used a split-panel CATI survey,

---

[3]The American Community Survey (ACS) is a new, continuous Census Bureau survey program intended to replace the decennial census "long form," and to provide annual updates of detailed population and housing data throughout each decade. The ACS is conducted using three stages of data collection; the research described here evaluated an alternative design for the CATI nonresponse follow-up instrument. See the Census Bureau's ACS website – *www.census.gov/acs/www* – for detailed information about the ACS program.

with an RDD sample, to examine the impact of household-level screening questions versus a more traditional person-level question approach. The latter, as the label implies, uses strictly person-level questions to assess the characteristics of interest: "Does John have a disability?" "Is Susan covered by health insurance?" "Has Robert ever served on active duty in the U.S. armed forces?" In some applications the person-level approach has been perceived as both tedious and burdensome (Hess and Rothgeb, 1998; Hess, Rothgeb, and Zukerberg, 1997). The household-level screening approach, in contrast, starts by determining whether anyone in the household has the characteristic: "Has anyone in this household ever served on active duty in the U.S. armed forces?" Follow-up questions to determine who has the characteristic are asked only if the answer to the household screener is positive.

Hess et al. report that, as expected, the household screener questionnaire treatment resulted in interviews that were significantly shorter than the interviews conducted with the person-level questions. Interviewers gave more positive ratings of the household-level questionnaire version; most notably, they judged the household-level screening form to be a significantly better instrument to use when dealing with reluctant or unenthusiastic respondents. An examination of overall cooperation rates found no significant advantage for the household screener format. However, the difference in the refusal rate component of nonresponse by questionnaire treatment was significant – the household-level screener questionnaire treatment elicited a significantly lower rate of refusals (27% of all assigned cases) than did the all person-level treatment (32%). Once again, given respondents' blindness to the type of interview to be administered, and interviewers' clear awareness, Hess et al. conclude that this difference "suggests that interviewers invested less effort in persuasion [in person-level treatment cases], perhaps because they were less eager to conduct that type of interview" [p. 581].

## IV. Computer-Assisted Questionnaires – Problems and Potential for Interview Interactions

Couper (2002) summarizes the myriad ways in which new technologies – new forms of self-administration (e.g., audio CASI, IVR, web-based surveys), audio-visual capabilities, the increasing portability of laptop and hand-held computers, mobile/cellular telephones, etc. – are reshaping the survey research landscape. The fact that Couper barely touches on the computerization of survey questionnaires – arguably, the most important and fundamental technology-based change in survey research – may indicate the extent to which the field no longer considers computerized survey questionnaires to be "new" technology. The obvious benefits of computerized questionnaires, especially for administration in the field (e.g., no more struggling to stay on the appropriate interview "path;" no more missing entries), and data capture (e.g., immediate editing of extreme values; no more data keying), have led to their widespread adoption; virtually all of the U.S. government's major survey programs are now administered with automated questionnaires.

With regard to their ability to guide and support a satisfying social interaction, computerized survey questionnaires carry both problems and potential. Many observers (e.g., Groves, Berry, & Mathiowetz, 1980; Groves & Mathiowetz, 1984; House, 1985; Couper et al., 1997) have noted the "segmentation" problem inherent in CAI instruments – each question is presented separately, so the interviewer never sees the whole questionnaire (or even large parts of it) all at once, just individual screens. This can cause interviewers to lose track of the relevance of the individual questions to the interview's "big picture," and, similarly, the relationship of individual questions to each other. The absence of this information can damage an interviewer's ability to maintain a smooth, natural interview "flow." A related issue is the uncertain position of a question in the entire question sequence. Groves and Mathiowetz (1984) note that interviewers in a paper vs. CATI experiment reported that "using the paper version ... gave them greater confidence, that they knew exactly where in the questionnaire they were at any moment" [p. 362].

Another problem is the rigid control of the flow of the instrument by the computer, which severely restricts interviewers' ability to improvise when the unexpected occurs – for example, to make repairs to old answers when new information arises in unexpected places (Couper et al., 1997; Groves and Mathiowetz, 1984). The rigid control of the interview flow also forces interviewers to record an answer before asking the next question; in contrast, interviewers using paper-and-pencil questionnaires often begin asking the next question while still recording the response to the previous one (Groves and Mathiowetz, *op cit.*).

Fuchs (2002) notes an additional problem with computerized questionnaires, although one which can perhaps be ameliorated to some extent through increased training and familiarity with the computer, and the particular application of a survey instrument on the computer. This is the problem of the interviewer's divided attention. He or she must pay substantial attention to the computer, at the expense of attention to the respondent and to the interview interaction. Combined with segmentation, Fuchs finds that one negative impact of divided attention is that it often leads interviewers to ask questions that have already been answered.

The problems inherent in computer-assisted questionnaires, however, seem well counterbalanced by the potential for improved questionnaire design that computerization offers. Those benefits include: (1) The ability to employ with ease complex questionnaire structures which have been demonstrated to improve the interview process, and which present major administration difficulties in the absence of automation. Prime examples of such structures have been noted earlier – the topic-based format, and the use of household-level screening questions.[4] (2) The ability to better tailor the question sequence to questions of relevance to the respondent. Automated instruments can employ "deep" and complex logic to determine an appropriate questionnaire path, without adding burden to the interviewer, or increasing the likelihood of

[4]Moore's (1996) paper-and-pencil prototype topic-based questionnaire proved very difficult for his researcher/ interviewers to administer. Similarly, with regard to the household-level screening format, Hess et al. (2001) note that, in contrast to a paper-and-pencil questionnaire administration, "computer-assisted survey instruments ... enable a fairly smooth administration of a household-level design" [p. 575].

question sequence errors. Based on logical deduction, or simply on information provided at a distant point in the interview, more questions can be presented in verification format, and some can be eliminated altogether. (3) Increased flexibility. A questionnaire need no longer stick rigidly to a single strategy to obtain some particular information, but can instead tailor the task to an option best suited to the respondent's situation. (4) Better long-term "memory." Computerization vastly expands the possibilities for use of dependent interviewing procedures in longitudinal surveys. Interviewers have been shown to value its greater efficiency, and the fact that they don't have to repeat from scratch the collection of information obtained in previous interviews (Mathiowetz & McGonagle, 2000), a feature of non-dependent longitudinal interviewing that respondents often complain about (see, e.g., Polivka and Rothgeb, 1993). Respondents, in fact, seem to actively embrace dependent procedures (Pascale and Mayer, 2002).

Thus, the computerization of survey questionnaires presents new opportunities for improving the interactional aspects of a survey interview. Some (e.g., Schegloff, 2002) may argue that this strives after an unrealistic goal, and that a structured survey questionnaire can never be designed to anticipate all of the problems that will inevitably arise – "global or formulaic devices will not work as solutions to the management of interactional contingencies in interviews" [p. 155]. Those who hold this position tend to see interviewer training and empowerment as the only way out. Maynard and Schaeffer (2002), for example, argue that interviewers must be trained "to solve the problems that arise when instruments and standardization are insufficient" [p. 28]. (Lynch (2002) offers a similar perspective, as do Shober and Conrad (1997), and Suchman and Jordan (1990).) It may well be true that good questionnaire design can never hope to eliminate all of an interview's rough spots, and that interviewer training has a vital role to play. But it is also true, as anyone who takes a close look at a few typical questionnaires can attest, that there are always improvements that could be made so as to avoid actively contributing to the problem.

## V. SIPP Improvements and Testing Results

The Census Bureau launched a research program in the late 1990s to improve the core questionnaires for the Survey of Income and Program Participation (SIPP). The primary goal of this program was to produce a less burdensome SIPP questionnaire that, without causing harm to the survey's important estimates, would encourage greater cooperation with the survey.

### A. Background: The SIPP Program and its Nonresponse Problems

SIPP is a nationally-representative, interviewer-administered, longitudinal survey conducted by the U.S. Census Bureau. Since its inception in 1983, SIPP has provided data on income, wealth, and poverty in the United States, the dynamics of program participation, and the effects of government programs on families and individuals. A SIPP panel consists of multiple waves (or rounds) of interviewing, with waves administered at four month intervals, typically over a period of three or four years. Since 1996 all SIPP interviews have been conducted with a computer-assisted questionnaire. Early interview waves are administered in-person; later waves are

generally conducted by telephone. The SIPP core instrument (which contains the survey content that is repeated in every survey wave) is detailed, long, and complex, collecting information about household structure, labor force participation, income sources and amounts, educational attainment, school enrollment, and health insurance over the prior four-month period. A typical SIPP interview takes about 30 minutes per interviewed adult (anyone age 15 or older). See U.S. Census Bureau (2001) for a more complete description of the SIPP program.

As with other government demographic surveys (Bates and Morgan, 2002), SIPP's nonresponse levels rose noticeably in the late 1990s, following – perhaps coincidentally – the survey's conversion from a paper-and-pencil questionnaire to computer automation. The primary change was evident in later waves of the 1996 panel. Substantial attrition continued to occur deep into the 1996 panel, long after attrition in previous panels had trickled almost to a standstill (see Eargle, 2000). In addition, SIPP's nonresponse trajectory seemed quite resistant to change; concerted efforts to increase cooperation with the survey by modifying field procedures (e.g., changing the reward structure for interviewers, adding respondent incentives, revising the advance letter, extending the period for completing interviews, etc.) yielded modest payoffs at best. The nonresponse increase, and the apparent absence of a "magic bullet" in traditional procedural fixes, coincided with a renewed conviction on the part of Census Bureau methodologists that the SIPP questionnaire made the interview more burdensome than it needed to be. These joint concerns led to a redesign program focused on improving the SIPP core instrument. The primary goal was to make the SIPP interview more efficient and less tedious, to develop clearer and "friendlier" wording, to simplify response tasks, and in general to make it a more pleasant experience for all participants – without harming data quality, of course. In addition to the inherent benefits of such "interview process" improvements, it was also hoped that they would improve survey cooperation. The resulting research program, the SIPP Methods Panel Project, developed and refined a large set of improvements to the core SIPP questionnaire, and evaluated those improvements in a series of three split-panel field experiments. See Moore et al. (2004) for a detailed description of the SIPP Methods Panel Project's design and results.

B. Exploiting Computerization to Improve the Interview Process

Briggs (1986) speaks of the "communicative blunders" [p. xiii] to which an interview can easily – and perhaps will inevitably – fall prey. Many others (e.g., Lynch, 2002; Maynard and Schaeffer, 2002; Shober and Conrad, 1997; Suchman and Jordan, 1990) have also noted the impossibility of designing a survey instrument which anticipates all of the problems which will unfold in a survey interview interaction. Recognizing these limits, we sought a modest goal: not to anticipate and avoid all problems, but rather to minimize their frequency by avoiding "hard-wiring" them into the questionnaire script as much as possible. In doing so, we also remained mindful of House's (1985) counsel: "The full potential of CATI data collection can be achieved only with a questionnaire that exploits the power of the system" [p. 213]. Thus, while we implemented many non-technology-based improvements, the primary focus was on interview process improvements which have become feasible only with the advent of computer-assisted questionnaires.

C. SIPP's Technology-Based Questionnaire Improvements

Some of the major new features which survived the Methods Panel's testing and refinement process, and which are now incorporated into the 2004 panel SIPP instrument, are as follows:

- The rostering process to identify household members used to proceed in a piecemeal, one-name-at-a-time fashion. The listing of household members was interrupted with questions after each reported name, which were themselves followed by a recurring "anyone else?" probe. In the new instrument, this process has been recast as a "tell me everyone who lives here" type of task. The roster screen accepts multiple names, and also allows for the capture of volunteered information regarding the sex of each person listed, and their relationship to the respondent. Thus, a respondent who replies to the opening question by saying (as respondents often do) something on the order of, "Well, there's me, Ann Smith, and my husband Robert, and our sons Malcolm and Jamal," has, in a single response, provided answers to multiple other questions in the survey, which therefore do not need to be asked. If the initial respondent turns out not to be the household's actual "reference person," the individual who serves as the official focal point for the household's relationship data, behind-the-scenes logic reassigns relationship codes appropriately, to the maximum extent possible. And if sex and relationship information is not produced spontaneously at this point, those data are captured later in the traditional way, with standard scripted questions.

- The new instrument uses a topic-based design to capture much of the basic demographic information about household members, in place of the strictly person-based design of the old instrument. (See section III.)

- In contrast to the strictly person-level nature of the old questionnaire design, the new instrument uses household-level screening questions to start the information capture process for several variables of interest (e.g., does everyone you have mentioned usually live here? has anyone ever served in the armed forces? was everyone born in the United States? does anyone speak a language other than English at home?), and only follows up at the level of the individual as necessary. (See section III.)

- The new instrument uses behind-the-scenes logic to avoid asking some questions whose answers are in essence completely determined by prior information. For example, if the race of each of a child's biological parents is the same, the instrument automatically assigns that same value to the child without asking a question overtly. The same procedures avoid unnecessary questions about the Spanish/Hispanic/Latino origin of biological children of parents who report the same origin.

- For some questions about things that either don't change, or that change predictably, the new procedures call for a single verification in the second interview wave to make sure that the information is correct. If the information is successfully verified, then no further

questions are asked in subsequent waves. For example, once a person has verified his or her age in the wave 2 interview (based on the date of birth established in wave 1), no further age verification questions are asked for that person for the remainder of the panel.

– The new instrument makes logical assumptions about marital status across interview waves. For example, if a person is divorced, widowed, or never married, and living alone (or living only with people to whom he/she couldn't possibly be married – e.g., parents, children, or siblings) in wave 1, and the household composition remains the same in the next wave, no question will be asked about marital status.

– In the old instrument a person's level of educational attainment was verified each wave, without considering the fact that the person may have completed his or her education decades before the interview. The new instrument only updates attainment each wave for people under age 25, and thereafter only if they were enrolled in school during the previous interview wave (or if they report new enrollment in the current wave).

– The new instrument makes use of general screening procedures to avoid asking large numbers of questions. There are two major examples of this sort of efficiency. First are new procedures which divert the wealthy around a long, detailed series of questions about need-based transfer programs, for which they are obviously not eligible. Likewise, the asset-poor are diverted around individual questions about extremely rare and esoteric types of assets (e.g., municipal or corporate bonds, U.S. government securities). The old procedures made no such distinctions, and administered all program and asset questions to all adults.

– The reporting period for earnings amounts (monthly) and asset income (4-month totals) is no longer dictated by the questionnaire, as has been the case in the past. The new instrument offers respondents an array of choices, and suggests that they pick the one most convenient and appropriate for their circumstances.

– After the first interview wave, prior wave income amount reports are available for use in dependent follow-up questions designed to overcome an initial nonresponse – or, as is often the case in SIPP, a response on the order of, "Whatever I said last time." If the data were reported in the previous interview (and if the previous interview's respondent consented to their being revealed in the subsequent wave), dependent questions assist respondents in reporting accurately: "Things may have changed, but I have recorded from the last interview that [asset type] produced about [X] dollars per [time period]. Does that still sound about right?" SIPP's procedures in the past never included any use of amount information in any sort of dependent question.

– Dependent interviewing procedures are in general used much more extensively throughout the new instrument (again, assuming the previous respondent has consented to his/her answers being revealed), most notably in questions about health insurance,

asset ownership, school enrollment, and financial aid for education. E.g., "You [had health insurance / owned [asset type] / were enrolled in school / received [financial aid for education]] last time we talked. Is that still the case?"

– The new instrument substantially reduces question repetition in getting a complete picture of a household's health insurance status. For example, it uses background information regarding age, relationship, the order in which household members are interviewed, and the results of earlier interviews to determine whether it is necessary to ask 15-to-21-year-olds (or their parents) about coverage under the SCHIP program. The new instrument also makes use of information obtained from policy-holders concerning dependents on their policy – and from dependents concerning policy-holders – to avoid asking unnecessary questions of subsequent respondents.

D. Testing the New Instrument Procedures

The Methods Panel project employed a series of three field experiments to test, refine, and evaluate these and many other new instrument procedures prior to their implementation in the 2004 SIPP production instrument. Each experiment consisted of approximately 2000 eligible addresses, with random assignment of sample cases to test and control groups – the control group being the current production SIPP questionnaire. Interviews in each Methods Panel field experiment were administered by personal visit, and were conducted by experienced SIPP interviewers, trained on methods for gaining respondent cooperation and on SIPP-specific goals and concepts. In each field test, interviewers' assignments included a mix of both control and test cases, and the questionnaire treatment to be used in a particular case was apparent to interviewers in their assignment materials.

We refer to the three experiments by the year in which they were conducted – MP2000, MP2001, and MP2002. The control instrument remained virtually constant across the three experiments[5]. The same is not at all true of the test instrument, which evolved considerably over the test series as, following each experiment, Methods Panel staff corrected bugs, made refinements, and modified or eliminated new procedures which gave evidence of causing harm to the survey's key data. MP2000 consisted of only a wave 1 interview; both MP2001 and MP2002 included a wave 2 interview four months after the wave 1 interview. See Doyle, Martin, and Moore (2000) for a detailed description of the design of the field experiments.

E. Results of the Field Experiments

1. Interviewer evaluations

[5]The 2001 production SIPP instrument changed in a few very minor ways from the 2000 panel instrument; thus the MP2000 and MP2001 control instruments differed slightly. But in essence the control instruments throughout the MP test series were identical – and identical in fact across MP2001 and MP2002.

Debriefing questionnaires were the primary means for assessing interviewers' reactions to the control and test instruments. In general, interviewers reacted quite positively to most of the test instrument's new features, and increasingly so over the three experiments. One component of each of the debriefing questionnaires sought interviewers' assessments of the specific question wording experiments included in the wave 1 instrument. In MP2000, interviewers rated the test instrument to be significantly superior to the control on 52% of the 67 debriefing questionnaire items evaluating specific experiments (versus 34% for the test; 13% no difference); in MP2001 and MP2002 the test instrument received significantly superior ratings on 73% and 93%, respectively, of those items. Other debriefing items assessed the more general performance of the two wave 1 instruments, and these show an even more dramatic pattern. Interviewers rated the test instrument superior on none (0%) of the 11 general evaluation[6] items in MP2000 (the control instrument scored higher on all items), a pattern which completely reversed itself over the two subsequent experiments. In MP2001 the test instrument elicited higher ratings on 91% of the general evaluation items, and in MP2002 the test instrument was judged superior on all items (100%). With regard to the new wave 2 instrument, first introduced in MP2001, interviewers' attitudes were substantially favorable from the beginning, and remained so through MP2002 as well. Interviewers also expressed highly positive attitudes toward most features of the new instrument design in several in-person debriefing sessions.

We attribute the emphatic shift toward more positive attitudes from MP2000 to MP2001 and MP2002 to both instrument refinements and increased interviewer familiarity. The latter notion finds some support in the debriefing questionnaire results, which reveal a tendency for interviewers with more SIPP experience to display a greater preference for the control instrument, and vice-versa. It also surfaced in the in-person debriefing sessions. Interviewers often commented on the necessary re-learning process that they had to endure, and their initial reluctance to abandon question formats and procedures which had become almost second nature to them over the years.

2. Unit (and other) nonresponse results

The goal of the SIPP improvements was not just to make the interview experience more positive – we hoped that improving the interview process would yield improved cooperation with the survey. Table 1 shows the wave 1 and 2 unit nonresponse outcomes for each experiment, by instrument treatment. The nonresponse rates are constructed in accord with standard practice (AAPOR, 2000); they show the proportion of eligible, occupied housing units which failed to respond to the wave 1 interview, and, for wave 2, the proportion of wave 1 interviewed households which failed to respond to the wave 2 interview.

[6]The "general performance" measures included such items such as ratings of the test and control instruments overall on their ease of administration, efficiency/inefficiency, flows/does not flow smoothly, repetitious/not repetitious, etc., as well as ratings as to how well each type of instrument worked in various circumstances, such as large households, small households, poor households, wealthy households, households with reluctant respondents, etc.

[Table 1 here]

On the surface, there is little in these results to suggest a positive impact of the new, more "interaction-friendly" SIPP instrument procedures on cooperation with the survey; chi-square tests find none of the comparisons to be statistically significant. However, a closer look at the data – particularly the wave 1 results across the three field experiment iterations – reveals an interesting trend, and suggests a different conclusion. A logistic regression analysis finds a significant nonresponse main effect for year of administration, indicating a steady decline in overall nonresponse, regardless of instrument treatment, from MP2000, to MP2001, to MP2002. This analysis also yields a borderline significant (p=.11) year*treatment interaction. Separate analyses of the control and test instrument treatment groups reveal the essential nature of the interaction – the decline in nonresponse across experimental iterations is substantial and significant in the test treatment; there are no significant differences in the control treatment. By observation, the wave 2 results suggest a similar sort of interaction, although the logistic analysis finds no significant effect of questionnaire treatment on unit nonresponse.

Although the wave 2 results at the unit (household) level fail to show the expected questionnaire treatment effects, the predicted trend does appear in the interview outcomes for individuals. As shown in Table 2, the test instrument treatment in the MP2002 experiment resulted in a significantly lower rate of person nonresponse in the wave 2 interview than the control. (In MP2001 the observed difference is in the same direction but falls short of statistical significance.) Thus, while we find it in a slightly different place than expected, there is also evidence in the wave 2 results suggesting that instrument design can have a significant impact on survey cooperation. Again, the impact was expected to be particularly pronounced in wave 2, since after the initial survey wave both interviewers and respondents have a basis for anticipating the likely quality of the survey interaction in the next wave.

[Table 2 here]

## VI. Conclusions and Discussion

A. The Methods Panel.....

The Methods Panel's wave 1 unit nonresponse results, considered as a group, and combined with the wave 2 person noninterview results, offer some support for the hypothesis that a questionnaire that helps sustain a better interview interaction will also yield improved survey cooperation. Increasing interviewer familiarity over the course of the three experiments, and steadily improving instrument design, provide plausible explanations for the fact that the evidence for the positive impact in wave 1 only surfaces in the trend of the results over the series of tests. Recall that, by design, all of the interviewers for the field experiments were trained and experienced SIPP interviewers – trained and experienced, in other words, on the control instrument. The survey methods literature has consistently linked inexperience with reduced

response rates (e.g., Couper and Groves, 1992), so it is hardly surprising, in retrospect, that at the beginning of the field test series, interviewers' relative lack of familiarity with the test instrument would have exerted a negative counterbalance to any interview process improvement effects. And also not surprising that across the second and third experiments, as that negative counterbalance was reduced (over 80 percent of the MP2002 interviewers also worked on one or both of the previous field tests), the impact of the improved instrument on cooperation gradually began to emerge.

Also recall that each successive iteration of the test instrument was an improvement over the previous version, as bugs were fixed, poorly-performing instrument features excised, and promising but flawed procedures refined. Thus another negative counterbalance was steadily reduced over the test series. So with two important caveats, the Methods Panel experiments with the SIPP questionnaire are consistent with the limited other work which suggests that instrument design improvements can exert a significant positive impact on cooperation, even in an initial contact interview. The caveats are that there may be some trial-and-error in the development of an improved design, and that the impact of the design improvements may not become apparent until interviewers become familiar and comfortable with them.

The failure to observe a significant impact of instrument design on wave 2 cooperation at the level of the sample unit is counter to initial hypotheses, but the significant impact observed in the wave 2 person-level non-interview outcomes suggests that some re-thinking may be in order. Unit nonresponse may be the most reasonable place to expect to see nonresponse effects in wave 1, when only the interviewer is aware of the nature of the to-be-administered questionnaire. However, once individual respondents' knowledge and experience with the wave 1 interview can be factored into the wave 2 response/nonresponse decision calculus, it seems quite logical to find impacts not at the unit level, but at the level of the individual respondent.

B. ... and Beyond

Labaw (1980) presents a strong moral case for devoting effort to questionnaire design that will improve the interview experience for respondents, condemning survey designers who fail to treat respondents "as people whose time is every bit as valuable as ours" [p 41]. Her argument also has a practical side, however, and one which raises the same issues that motivated the current inquiry. The viability of the survey research enterprise depends on the cooperation of the public, Labaw contends; thus "it is in the researcher's own interest to design a questionnaire that does not abuse respondents through excessive length or stupidly contrived, meaningless questions" [p 41]. The essence of this argument is that unless survey designers learn to "play nice," respondents will eventually take their bat and ball and go home.

Recent developments in survey research have demonstrated that "playing nice" involves more than simply avoiding the obvious abuses that Labaw cites, and the gross insults that are noted in

professional codes of conduct, such as AAPOR's[7]. It also means being attuned to more subtle aspects of questionnaire design. A major development in this regard over the last two decades has been the application of the tools and perspectives of linguistics to the interactions of interviewers and respondents as they negotiate the treacherous waters of administering and responding to a standardized, scripted survey questionnaire. "Treacherous waters" because, as this linguistic perspective has revealed, interview participants almost inevitably must put forth some effort to balance conflicts between conforming to the rules of the interview and maintaining a sensible interaction.

To the extent that a survey interaction is beset with "communicative blunders" (Briggs, 1986), or "interactional troubles" (Suchman and Jordan, 1990), it is likely to be less satisfying to both participants, and less likely to be an interaction either party would be eager to endure or repeat. The present study adds to a small body of research which offers some support for the notion that "interaction-friendly" questionnaire design is important, and can yield improved survey cooperation. It is likely that this has always been the case. But now there is a new feature on the survey research landscape – the use of computers in the administration of survey questionnaires – which has opened up new possibilities for making the interview experience a more pleasant one for all participants, while still ensuring adherence to the notion of scripted/controlled question wording and sequencing, the cornerstone of scientific survey methodology. Questionnaire design should try to be supportive of a reasonable social interaction; automation can be exploited to help; improved cooperation can result.

A more interaction-friendly computerized questionnaire, however, can be a mixed blessing. This is especially so if the improvements add greatly to the complexity of the software. Increased complexity brings an increased risk of programming errors, and makes instrument testing and debugging more difficult and time-consuming. These costs must be weighed against the benefits. If the results of the present study are generalizable, the benefits of improved questionnaire design may not be large. Thus, questionnaire improvements, by themselves, may only carry the ball so far. Reaping more substantial benefits from improved survey interactions may require the active assistance of interviewers, who would need new skills beyond reading questions exactly as worded. They would need to be trained in detecting and even anticipating interaction "strains," and trained in how to manage them effectively, to the benefit of both the survey interaction and the quality of the data that interaction is intended to produce. Shober and Conrad (1997) offer a glimpse of the sorts of new skills interviewers would need to possess in order to be of real assistance.

---

[7]Other than offering a strong confidentiality pledge, the AAPOR "Code of Professional Ethics and Practices" (see *www.aapor.org*) refers to appropriate dealings with respondents only as follows: "We shall strive to avoid the use of practices or methods that may harm, humiliate, or seriously mislead survey respondents." Compliance only requires that grossly negative treatment of respondents be avoided, in other words – not that positive treatment be sought.

**Table 1:**
**Unit Response and Nonresponse Outcomes in the Methods Panel Field Tests by Instrument Treatment**

| Methods Panel Field Test | | | Instrument Treatment Group | | Production SIPP |
|---|---|---|---|---|---|
| | | | Control | Test | |
| **MP2000**: (Aug-Sep)[1/] | a. | **WAVE 1** Eligible units | 988 | 1032 | 13,141 |
| | b. | Interviews[2/] | 842 | 854 | 11,660 |
| | c. | Refusal Non-Interviews | 93 | 105 | 1,108 |
| | d. | Other Non-Interviews | 53 | 73 | 373 |
| | Unit Nonresponse Rate [(c+d)/a]*100 | | 14.8% | 17.2% | 11.3% |
| | WAVE 2 [n/a] | | [n/a] | [n/a] | – |
| **MP2001**: (Jun-Jul)[1/] | a. | **WAVE 1** Eligible units | 1119 | 1041 | 40,489 |
| | b. | Interviews[2/] | 949 | 870 | 35,102 |
| | c. | Refusal Non-Interviews | 110 | 113 | 4,033 |
| | d. | Other Non-Interviews | 60 | 58 | 1,354 |
| | Unit Nonresponse Rate [(c+d)/a]*100 | | 15.2% | 16.4% | 13.3% |
| (Oct-Nov)[1/] | a. | **WAVE 2** Eligible units | 961 | 878 | 30,514 [4/] |
| | b. | Interviews[2/] | 887 | 799 | 28,086 |
| | c. | Refusal Non-Interviews | 27 | 33 | 1,115 |
| | d. | Mover Non-Interviews[3/] | 23 | 19 | 522 |
| | e. | Other Non-Interviews | 24 | 27 | 791 |
| | Unit Nonresponse Rate [(c+d+e)/a]*100 | | 7.7% | 9.0% | 8.0% |
| **MP2002**: (Jul-Aug) | a. | **WAVE 1** Eligible units | 1360 | 1346 | [n/a; there was no production SIPP W1 or W2 interviewing in 2002] |
| | b. | Interviews[2/] | 1181 | 1182 | |
| | c. | Refusal Non-Interviews | 110 | 98 | |
| | d. | Other Non-Interviews | 69 | 66 | |
| | Unit Nonresponse Rate [(c+d)/a]*100 | | 13.2% | 12.2% | |
| (Nov-Dec) | a. | **WAVE 2** Eligible units | 1197 | 1207 | |
| | b. | Interviews[2/] | 1094 | 1117 | |
| | c. | Refusal Non-Interviews | 40 | 42 | |
| | d. | Mover Non-Interviews[3/] | 31 | 20 | |
| | e. | Other Non-Interviews | 32 | 28 | |
| | Unit Nonresponse Rate [(c+d+e)/a]*100 | | 8.6% | 7.5% | |

Summary of logistic regression results (details available upon request) –
Wave 1 unit nonresponse: for the year (2000-2002) * treatment interaction, $\beta = -0.076$, $\chi^2 = 2.50$, $p = .11$
Wave 2 unit nonresponse: n.s.

Notes:
[1/] Production SIPP wave 1 interviews are conducted in February through May; wave 2 in June through September.
[2/] Includes "sufficient partial" interviews.
[3/] In production SIPP, virtually all wave 2 "mover" households are followed and an interview is attempted at their new address; not locatable cases become "mover non-interviews." In contrast, the MP2001 and MP2002 experiments only attempted to follow and interview local movers, thus increasing the likelihood of a mover non-interview outcome.
[4/] The number of eligible units in wave 2 of the 2001 SIPP panel is substantially lower than expected (given the number of completed wave 1 interviews) due to a 15% sample reduction, necessitated by a budget shortfall.

**Table 2:**

**Wave 2 Person Non-Interviews in the Methods Panel Field Tests by Instrument Treatment**

| Methods Panel Field Test | | | Instrument Treatment Group | |
|---|---|---|---|---|
| | | | Control | Test |
| **MP2001**: | a. | Eligible wave 2 adults (age 15+) | 1914 | 1686 |
| | b. | Interviews | 1610 | 1434 |
| | c. | Non-Interviews [1/] | 304 | 252 |
| | Person Non-interview Rate [(c/a)*100] | | 15.9% | 15.0% |
| | | | (n.s.) | |
| **MP2002**: | a. | Eligible wave 2 adults (age 15+) | 2394 | 2276 |
| | b. | Interviews | 2055 | 2051 |
| | c. | Non-Interviews [1/] | 339 | 225 |
| | Person Non-interview Rate [(c/a)*100] | | 14.2% | 9.9% |
| | | | $\chi^2 = 20.1$, $p<.0001$ | |

Notes:
[1/]Combines all forms of wave 2 person nonresponse, including that due to unit nonresponse, person nonresponse within otherwise interviewed households, and whole-household or individual mover non-interviews.


## Acknowledgments

I am indebted to Adam Carle for his invaluable assistance with regard to the logistic regression analysis summarized in Section V; Anna Chan provided crucial assistance with the person non-interview results, also presented in Section V. The paper has also benefitted from the constructive criticisms of several reviewers, including especially Karen Bogen, Adam Carle, Pat Doyle, and Yuling Pan; I am very grateful for their efforts. The reviewers are, of course, totally absolved of any responsibility for the paper's shortcomings, which remain solely my own.


## REFERENCES


American Association for Public Opinion Research (AAPOR). 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, Ann Arbor, Michigan: AAPOR.

Bates, N. and M. Morgan. 2002. "Noninterview Rates for Selected Major Demographic Household Surveys: 1990 – 2001." Unpublished report. Washington DC: U.S. Census Bureau, August 2, 2002.

Bogen, K. 1996. "The Effect of Questionnaire Length on Response Rates – A Review of the

Literature." *Proceedings* of the American Statistical Association, Survey Research Methods Section, 1020-1025.

Briggs, C. 1986. *Learning How to Ask.* Cambridge: Cambridge University Press.

Cannell, C., K. Marquis, and A. Laurent. 1977. "A Summary of Studies of Interviewing Methodology." Vital and Health Statistics, Series 2, No. 69, Washington, DC: U.S. Department of Health, Education, and Welfare (National Center for Health Statistics).

Clark, H. 1979. "Responding to Indirect Speech Acts." *Cognitive Psychology* 11:430-477.

Colosi, R. 2001. "An Analysis of Time Stamp Data: Person-based Interviewing vs. Topic-based Interviewing." Unpublished paper. Washington, DC: U.S. Census Bureau, December 10, 2001.

Converse, J. and H. Schuman. 1974. *Conversations at Random: Survey Research as Interviewers See It.* New York: Wiley.

Couper, M. 1997. "Survey Introductions and Data Quality." *Public Opinion Quarterly* 61:317-338.

Couper, M. 2002. "New Technologies and Survey Data Collection: Challenges and Opportunities." Paper presented at the International Conference on Improving Surveys, Copenhagen, August 2002.

Couper, M., M. Fuchs, S. Hansen, and P. Sparks. 1997. "CAPI Instrument Design for the Consumer Expenditure (CE) Quarterly Interview Survey." Final Report to the Bureau of Labor Statistics (Task Order #12, Contract J-9-J-5-0025), December 1997.

Couper, M. and R. Groves. 1992. "The Role of the Interviewer in Survey Participation." *Survey Methodology* 18:263-278.

Couper, M. and R. Groves. 2002. "Introductory Interactions in Telephone Surveys and Nonresponse." In Maynard, D., H. Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen (eds.). 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley & Sons, 161-177.

Doyle, P., E. Martin, and J. Moore. 2000. "Methods Panel to Improve Income Measurement in the Survey of Income and Program Participation." *Proceedings* of the American Statistical Association, Survey Research Methods Section, 953-958.

Eargle, J. 2000. "Sample Loss Rates through Wave 12 for the 1996 Panel." Unpublished

memorandum. Washington, DC: U.S. Census Bureau, May 23, 2000.

Fuchs, M. 2002. "The Impact of Technology on Interaction in Computer-Assisted Interviews." In Maynard, D., H. Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen (eds.). 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley & Sons, 471-491.

Grice, H. 1975. "Logic and Conversation." In P. Cole and J. Morgan (eds.). 1975. *Syntax and Semantics: Vol. 3, Speech Acts*. New York: Academic Press, 41-58.

Groves, R., M. Berry, and N. Mathiowetz. 1980. "Some Impacts of Computer Assisted Telephone Interviewing on Survey Methods." *Proceedings* of the American Statistical Association, Survey Research Methods Section, 519-524.

Groves, R., R. Cialdini, and M. Couper. 1992. "Understanding the Decision to Participate in a Survey." *Public Opinion Quarterly* 56:475-495.

Groves, R. and M. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley & Sons.

Groves, R., D. Dillman, J. Eltinge, and R. Little (eds.). 2002. *Survey Nonresponse*. New York: Wiley & Sons.

Groves, R. and N. Mathiowetz. 1984. "Computer Assisted Telephone Interviewing: Effect on Interviewers and Respondents." *Public Opinion Quarterly* 48:356-369.

Groves, R., K. McGonagle, and E. O'Brien. 1999. "Refusal Aversion Training in the Agricultural Census." Final report to the National Agricultural Statistics Service, September, 1999.

Hess, J., J. Moore, J. Pascale, J. Rothgeb, and C. Keeley. 2001. "The Effects of Person-Level versus Household-Level Questionnaire Design on Survey Estimates and Data Quality." *Public Opinion Quarterly* 65:574-584.

Hess, J. and J. Rothgeb. 1998. "Summary Report: 1998 SPD Taped Interviews." Unpublished report prepared for the SPD Steering Committee. Washington, DC: U.S. Census Bureau.

Hess, J., J. Rothgeb, and A. Zukerberg. 1997. "Survey of Program Dynamics: Pretest Evaluation Report." Unpublished report prepared for the SPD Steering Committee. Washington, DC: U.S. Census Bureau.

House, C. 1985. "Questionnaire Design with Computer Assisted Telephone Interviewing."

*Journal of Official Statistics* l:209-219.

Houtkoop-Steenstra, H. and H. van den Bergh. 2000. "Effects of Introductions in Large-Scale Telephone Survey Interviews." *Sociological Methods and Research* 28:281-300.

Jones, E. and R. Nisbett. 1971. *The Actor and the Observer: Divergent Perceptions of the Causes of Behavior*. Morristown, NJ: General Learning Press.

Kalton, G., J. Lepkowski, G. Montanari, and D. Maligalig. 1990. "Characteristics of Second Wave Nonrespondents in a Panel Survey." *Proceedings* of the American Statistical Association, Survey Research Methods Section, 462-467.

Krosnick, J., S. Narayan, and W. Smith. 1996. "Satisficing in Surveys: Initial Evidence." In M. Braverman and J. Slater (eds.). 1996. *Advances in Survey Research: New Directions for Evaluation #70*. San Francisco: Jossey-Bass, 29-44

Labaw, P. 1980. *Advanced Questionnaire Design*. Cambridge, MA: Abt Books.

Lepkowski, J. and M. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In Groves, R., D. Dillman, J. Eltinge, and R. Little (eds.). 2002. *Survey Nonresponse*. New York: Wiley & Sons, 259-272.

Lynch, M. 2002. "The Living Text: Written Instructions and Situated Actions in Telephone Surveys." In Maynard, D., H. Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen (eds.). 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley & Sons, 125-150.

Manis, M., S. Cornell, and J. Moore. 1974. "Transmission of Attitude-Relevant Information Through a Communication Chain." *Journal of Personality and Social Psychology* 30:81-94.

Mathiowetz, N. and K. McGonagle. 2000. "An Assessment of the Current State of Dependent Interviewing in Household Surveys." *Journal of Official Statistics* 16:401-418.

Mayer, T. and E. O'Brien. 2001. "Interviewer Refusal Aversion Training to Increase Survey Participation." *Proceedings* (on CD-ROM) of the American Statistical Association, Survey Research Methods Section.

Maynard, D., H. Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen (eds.). 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley & Sons.

Maynard, D. and N. Schaeffer. 2002. "Standardization and Its Discontents." In Maynard, D., H.

Houtkoop-Steenstra, N. Schaeffer, and J. van der Zouwen (eds.). 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley & Sons, 3-45.

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. Washington, DC: U.S. Census Bureau.

Moore, J. and L. Moyer. 2002. "ACS/CATI Person-Based/Topic-Based Field Experiment." Research Report Series, *Survey Methodology #2002-04*, Statistical Research Division, Washington, DC: U.S. Census Bureau.

Moore, J., J. Pascale, P. Doyle, A. Chan, and J. Griffiths. 2004. "Using Field Experiments to Improve Instrument Design: The SIPP Methods Panel Project." In Presser, S., J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer (eds.). 2004. *Methods for Testing and Evaluating Survey Questions*. New York: Wiley & Sons, 189-207.

Morton-Williams, J. 1991. "Obtaining Co-operation in Surveys – The Development of a Social Skills Approach to Interviewer Training in Introducing Surveys." Working Paper No. 3, London: Joint Centre for Survey Methods.

Morton-Williams, J. 1993. *Interviewing Approaches*. Aldershot, England: Dartmouth Publishing.

O'Brien, E., T. Mayer, R. Groves, and G. O'Neill. 2002. "Interviewer Training to Increase Survey Participation." *Proceedings* (on CD-ROM) of the American Statistical Association, Survey Research Methods Section.

Pascale, J. and T. Mayer. 2002. "Alternative Methods for Exploring Confidentiality Issues Related to Dependent Interviewing." Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods, Charleston, SC, November 14-17, 2002. Washington, DC: U.S. Census Bureau.

Payne, S. 1951. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.

Polivka, A. and J. Rothgeb. 1993. "Redesigning the CPS Questionnaire." *Monthly Labor Review* 116:10-28.

Schegloff, E. 2002. "Survey Interviews as Talk-in-Interaction." In Maynard, D., H. Houtkoop-

Steenstra, N. Schaeffer, and J. van der Zouwen (eds.). 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: Wiley & Sons, 151-157.

Shober, M. and F. Conrad. 1997. "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61:576-602.

Suchman, L. and B. Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85:232-253.

U.S. Census Bureau. 2001. "SIPP Users' Guide." Washington, DC: U.S. Census Bureau.