DOES CONVERSATIONAL INTERVIEWING REDUCE SURVEY MEASUREMENT ERROR?

MICHAEL F. SCHOBER
FREDERICK G. CONRAD

Abstract Standardized survey interviewing is widely advocated in order to reduce interviewer-related error, for example by Fowler and Mangione. But Suchman and Jordan argue that standardized wording may decrease response accuracy because it prevents the conversational flexibility that respondents need in order to understand questions as survey designers intended. We propose that the arguments for these competing positions—*standardized* versus *flexible* interviewing approaches—may be correct under different circumstances. In particular, both standardized and flexible interviewing should produce high levels of accuracy when respondents have no doubts about how concepts in a question map onto their circumstances. However, flexible interviewing should produce higher response accuracy in cases where respondents are unsure about these mappings. We demonstrate this in a laboratory experiment in which professional telephone interviewers, using either standardized or flexible interviewing techniques, asked respondents questions from three large government surveys. Respondents answered on the basis of fictional descriptions so that we could measure response accuracy. The two interviewing techniques led to virtually perfect accuracy when the concepts in the questions clearly mapped onto the fictional situations. When the mapping was less

MICHAEL F. SCHOBER is assistant professor in the Department of Psychology at the New School for Social Research. FREDERICK G. CONRAD is research psychologist at the Bureau of Labor Statistics. Some of these results were presented at the fifty-first annual meeting of the American Association for Public Opinion Research, Salt Lake City, Utah, May 1996, and at the sixth annual meeting of the Society for Text and Discourse, San Diego, California, July 1996. This work was supported by an American Statistical Association/National Science Foundation/Bureau of Labor Statistics Senior Research Fellowship to Michael Schober. We thank Clyde Tucker, Cathy Dippo, Sarah Jerstad, and Kimberly Clark for support in various aspects of the project, and Susan Brennan, Elizabeth Brooks, Herbert Clark, Jim Esposito, Brian Harris-Kojetin, Don Harrison, Barbara Kanki, and Candy Sidner for comments on earlier drafts. The views expressed here reflect the opinions of the authors and not those of the Bureau of Labor Statistics. Address correspondence to Michael F. Schober, Department of Psychology, AL-340, New School for Social Research, 65 Fifth Avenue, New York, NY 10003; E-mail: schober@newschool.edu.

Public Opinion Quarterly Volume 61:576–602 0033-362X/97/6104-0002$02.50

clear, flexible interviewing increased accuracy by almost 60 percent. This was true whether flexible respondents had requested help from interviewers or interviewers had intervened without being asked for help. But the improvement in accuracy came at a substantial cost—a large increase in interview duration. We propose that different circumstances may justify the use of either interviewing technique.

Introduction

In the typical survey interview, interchanges like this one sometimes occur:

> Interviewer (I): Last week, did you do any work for pay?
> Respondent (R): Well, that depends. What exactly do you mean by work?

The interviewer is now faced with a choice. Should she use her knowledge to answer the respondent's question, or should she leave the interpretation of "work" up to the respondent?

According to the prevailing philosophy of survey interviewing, *standardization,* the interviewer must leave the interpretation of the question up to the respondent. Interviewers must present exactly the same stimulus to all respondents, always reading exactly the same question and never interpreting the question in any way (Fowler 1991; Fowler and Mangione 1990). When a respondent solicits help, the interviewer should use "neutral probing" techniques: repeat the question, ask for the respondent's interpretation (e.g., saying "whatever it means to you" or "we need *your* interpretation of the question"), or present the response alternatives ("Would that be a yes or a no?"). The idea is that respondents can only be guaranteed to be answering the same question if the stimulus—the words uttered by the interviewer—is uniform from one interview to the next. This stimulus uniformity should reduce measurement error resulting from the interviewer.

In contrast, critics in the survey world (Suchman and Jordan 1990, 1991) and in other disciplines (e.g., Briggs 1986; Holstein and Gubrium 1995; Kvale 1994) would argue that the interviewer in our example should help the respondent and define "work for pay." The argument is that response validity is undermined if respondents interpret questions idiosyncratically. Indeed, substantial evidence in the survey world supports this contention (see, e.g., Belson's [1981, 1986] findings that even ordinary words like "weekend," "children," "you," and "generally" in surveys are interpreted in many different ways). As a remedy, these critics have proposed or indirectly endorsed *conversationally flexible* interviewing techniques. Interviewers should engage in something more like ordinary conversation, deviating from the standardized script to assure that respondents interpret questions consistently and correctly.

Proponents of flexible interviewing claim that it should lead to more accurate responses than standardized interviewing does (Suchman and Jordan 1990, 1991), but proponents of standardization argue just the opposite (Fowler 1991; Fowler and Mangione 1990). Who is correct? As Schaeffer (1991) puts it, "It is an open question whether systematically giving participants more access to 'normal' conversational resources would improve the quality of the interaction or the resulting data" (p. 371). This study is a first attempt at just such a systematic comparison.

The Debate

Proponents of standardization argue that the assumptions of flexibility ignore the history of what led to standardization in the first place. Standardized techniques were developed because of evidence that interviewers were influencing responses and because researchers needed greater statistical precision and more affordable ways to test large populations (see Beatty 1995). Proponents of standardization point out that the arguments for flexibility are based on a few extreme examples of flawed interactions (Kovar and Royston 1990). They suggest that it is poor question wording rather than standardized interaction that leads to problems, and these problems might be remedied with better question pretesting. In addition, problems due to respondents' misunderstanding may be rare enough to be unimportant in large-sample surveys.

In contrast, proponents of flexibility argue that rigid adherence to survey scripts jeopardizes validity (Suchman and Jordan 1990, p. 233) because there is no guarantee that all respondents will interpret questions the same way. Standardization, the argument goes, does not allow the "full resources of conversational interaction" that are necessary to assure consistent interpretations. This claim is consistent with a long tradition of empirical research, both in the laboratory and based on naturalistic observation, that shows how conversational interaction can be essential for understanding (see, e.g., Brennan 1990; Cicourel 1973; Clark 1992, 1996; Clark and Wilkes-Gibbs 1986; Goffman 1981; Goodwin 1981; Gumperz 1982; Krauss and Fussell 1996; Rogoff 1990; Sacks, Schegloff, and Jefferson 1974; Schegloff 1984; Schiffrin 1994; Schober and Clark 1989; Tannen 1989, among many others).

We agree with Schaeffer (1991) and Beatty (1995) that both positions have merit. But we propose that they are correct under different circumstances. Consider this question: "How many hours per week do you usually work?" A respondent who has a nine-to-five job with no overtime probably will not need any clarification to interpret this question as intended. In this case standardized interviewing should promote accurate

responses.[1] But a respondent who works as a freelance writer, and whose lunches with editors and ruminations while jogging might legitimately be considered work, may be unsure what "work" means in this question. In this case flexible interviewers might help the respondent interpret "work" as the question author intended, thus promoting more accurate responses.

Note that the question itself is not ordinarily ambiguous—"How many hours per week do you usually work?" is, in fact, a commonly used, well-pretested survey question, and the words in the question should be familiar to native speakers of the language. What is ambiguous is the way "work" and the respondent's circumstances correspond. This depends on how the survey organization defines what counts as work and what does not—and this definition may differ from how most respondents ordinarily understand "work."[2] We call these ambiguous correspondences between questions and situations "complicated mappings."

Given this line of reasoning, we predict that standardized interviewing should lead to accurate responding when the concepts in a question map onto a respondent's life circumstances in a straightforward way. In contrast, when the mapping is complicated, more respondents should be able to answer the question as intended if interviewers can clarify official definitions, thus improving overall response accuracy.

Measuring Accuracy

Gathering empirical evidence on how standardized and flexible interviewing affect accuracy is particularly important since "both positions are held more on theoretical than empirical grounds" (Fowler 1991, p. 269). But gathering such evidence is also particularly difficult to do (van der Zouwen, Dijkstra, and Smit 1991; Wentland 1993). Comparing responses with official records or personal diaries is expensive, and there is no way to ensure that official records or diaries are correct, if they are even available.

Instead, researchers have relied on surrogate measures. For example, Hyman (1954) counted interviewers' probes under different interviewing

1. Of course, respondents can answer inaccurately for many other reasons: memory errors, estimation errors, etc. But understanding questions appropriately is a prerequisite for accurate responding.

2. Surveys do not always have official definitions for key concepts; this is especially likely for onetime ad hoc surveys. And when they do, the definitions are not always consistent. Definitions can even differ between surveys administered by the same organization. One major U.S. government agency, for example, conducts two surveys that ask about half bathrooms; one survey defines half bathrooms as having either one or two fixtures (toilet, sink, or shower/tub), and the other requires two fixtures.

techniques, under the assumption that interviewers probe less when respondents understand questions better. Fowler's (1991) approach is to quantify how much responses vary for different interviewers, on the assumption that less "interviewer-related error" reflects greater response accuracy. But there is no guarantee that either of these surrogate measures truly captures respondents' understanding or accuracy.

The purpose of our study was to directly compare response accuracy under standardized and flexible interviewing techniques. So that we could assess accuracy of the data with confidence, we had respondents answer questions on the basis of fictional scenarios we designed, rather than asking them about their own lives. We used real questions from major government surveys. In all cases, the sponsoring organizations had published explicit definitions for the concepts in the questions. Thus we knew the correct (with respect to these definitions) answer for all question-scenario combinations and we could easily determine respondents' accuracy.

We implemented standardization following Fowler and Mangione's (1990) prescriptions. These require interviewers to use neutral probes to help responses match the question's objectives, which must be inferred from the official definitions. Interviewers must avoid influencing responses in any way, including presenting definitions to respondents. Interviewers should nonetheless be taught the definitions so that they can judge the completeness of the respondents' answers (see Beatty 1995).

We recognize that this version of standardization is not practiced universally. Some organizations that subscribe to the theory of standardization allow interviewers to provide scripted definitions on request. According to Fowler and Mangione (1990, p. 21), such practices are not standardized: not every respondent is presented with the same stimulus, and interviewers are not guaranteed to present definitions consistently. The more stringent version we tested was clearly standardized.

We saw several alternatives for implementing flexible interviewing. In order of increasing departure from standardization, the possibilities are as follows. (1a) Interviewers could read scripted definitions for concepts in the questions, but only when respondents explicitly request them. (1b) Interviewers could provide customized (unscripted) definitions only at the respondent's request. (2) In addition to providing definitions on demand in (1), interviewers could resolve confusions and clarify concepts whenever they judge it necessary, even if the respondent does not request help. Interviewers could do this either with scripted definitions and probes (2a) or by improvising (2b). (3) In addition to providing clarifications on demand (1) and voluntarily (2), interviewers could initially present the question in their own words, based on their own understanding of the survey designers' intentions. In this study, we used (2b), because it involves the greatest degree of flexibility while preserving initial question wording.

These variations of flexible interviewing techniques differ from more radical proposals, for example, that interviewers should engage in unscripted interactions in which respondents help set the research agenda itself (see, e.g., Mishler 1986). The conversationally flexible alternatives (1)–(3) are designed to promote consistent interpretations of the questions in the surveys, and in this respect they share the goals of standardized interviewing. Where they differ is in what leads to consistent interpretations.

We are aware that in practice many survey interviewers probably use some combination of standardized and flexible interviewing. Here we used a pure version of each technique so that we could directly evaluate the competing theories.

Method

Design. Different respondents participated in either standardized or flexible interviews. In each interview, the interviewer asked 12 questions about fictional scenarios; respondents were not answering about their own lives but about these fictional circumstances. All respondents were asked the same questions; what differed was the type of interaction in the interview.

Participants. The 43 "respondents" were experimental subjects, all fluent speakers of English. Forty subjects were recruited from advertisements in the *Washington Post* and paid $25 each, and three were volunteers from the Bureau of Labor Statistics (BLS) staff (not survey professionals). In selecting respondents for standardized or flexible interviews, we roughly balanced gender and education level. All respondents had at least a high school diploma. Of the 21 respondents in the standardized interviews, 11 were women and 10 were men; 3 had completed high school only, 6 were current college students, and 12 had completed college. Of the 22 respondents in the flexible interviews, 11 were women and 11 were men; 6 had completed high school only, 4 were current college students, and 12 had completed college.

In all other respects, respondents were assigned to interviewing conditions arbitrarily. Of the 21 respondents in the standardized interviews, 5 were black and 16 were white; their average age was 32.2 years, ranging from 18 to 70 years. Of the 22 respondents in the flexible interviews, 9 were black, 10 were white, and 3 were Asian; their average age was 35.5 years, ranging from 18 to 59 years.

The 22 interviewers (21 white and 1 black) were professional Census Bureau interviewers; all but one had at least 30 months of experience at the Hagerstown, Maryland, Census Bureau telephone facility. The 11 standardized interviewers (10 women, 1 man) averaged 43 months of experience, ranging from 5 to 81 months. The 11 flexible interviewers (7

women, 4 men) averaged 69 months of experience, ranging from 36 to 100 months.

Each interviewer, calling from the Hagerstown telephone facility, called two respondents who had come into the BLS laboratory (except for one interviewer who only called one respondent), for a total of 43 interviews.

Questions. Four of the 12 questions were about employment, selected from the Current Population Survey (CPS); four were about housing, selected from the Consumer Price Index Housing survey (CPI Housing); and four were about retail purchases, selected from the Current Point of Purchase Survey (CPOPS), which is part of the Consumer Price Index program. All questions had been pretested, most extensively those from the CPS. Some of the questions were familiar to the interviewers, who averaged 53 months of experience administering the CPS (40 and 66 months for standardized and flexible interviewers, respectively). The CPOPS questions were familiar to only five of the 22 interviewers (one standardized and four flexible); no interviewers had administered the CPI Housing survey.

We modified some questions slightly to include the proper names of people described in the scenarios, so that the questions would be about "Carla" or "Harry" rather than about the actual respondent. For example, one employment question was, "Last week, did Pat have more than one job, including part-time, evening or weekend work?" One housing question was, "How many other rooms are there, other than bedrooms and bathrooms?" One purchasing question was, "Has Kelly purchased or had expenses for household furniture?" (appendix A includes all 12 questions.)

The key concepts in all questions were officially defined by the sponsoring survey programs. These published definitions are among the training materials provided to interviewers in the actual government surveys. For example, "household furniture" is defined as "tables, chairs, footstools, sofas, china cabinets, utility carts, bars, room dividers, bookcases, desks, beds, mattresses, box springs, chests of drawers, night tables, wardrobes, and unfinished furniture. Do not include TV, radio, and other sound equipment, lamps and lighting fixtures, outdoor furniture, infants' furniture, or appliances." For the complete set of definitions used in this study, see appendix A.

The questions were always asked in the same relative order as in the actual survey instruments from which they were drawn. For example, in the actual CPOPS instrument our fourth purchasing question appears after our third purchasing question, even though the CPOPS instrument contains intervening questions that we did not ask. Some actual CPS respondents might not be asked all four of our employment questions, depending on their answers to previous questions.

We counterbalanced the order in which questions from a particular domain (housing, work, or purchases) were asked. There were six possible sequences for the three domains; for every interviewer (each of whom interviewed two respondents), two of the six orderings were chosen at random, without replacement. This was done to minimize any effects of domain order and to assure that the orders were used equally often.

Scenarios. The fictional scenarios on which respondents based their answers included floor plans, purchase receipts, and descriptive texts. The scenarios were designed so that the content of one had nothing to do with the content of the others. Each floor plan was from a different imaginary house or apartment; each purchase was made by a different person from a different establishment; each work situation was about a different person.

These scenarios were available to the respondents both before and during the interviews. However, they were never available to interviewers and so the interviewers never knew the correct answer. Interviewers could also never predict correct answers from any interview they had previously conducted because respondents interviewed by a given interviewer were always presented with different versions of the scenarios (see the next section). The way the knowledge was allocated to the participants was therefore analogous to its division in an actual survey. Respondents knew the "facts," and interviewers knew the questions and concepts.

Mappings. There were two versions of each scenario, one that corresponded to the concepts in the question in a straightforward way (a straightforward mapping) and one that corresponded to the concepts in the question less clearly (a complicated mapping). In the "household furniture" example, the scenario that led to a straightforward mapping was a purchase receipt for an end table. In contrast, the scenario that led to a complicated mapping was a purchase receipt for a floor lamp, which the official definition excludes as a piece of furniture (the complete set of scenarios is available from the authors).

For each respondent, two of the four scenarios in each domain (housing, work, purchases) had a straightforward mapping to their respective questions and two had a complicated mapping. Each respondent had a different sequence of mappings for each of the three domains, so that respondents could not predict complexity of the mappings. Also, interviewers could not predict the complexity of the mappings, because the two respondents assigned to each interviewer had complementary mappings. For example, the mappings for one respondent in the housing domain might have been complicated, straightforward, straightforward, and complicated; the second respondent for the same interviewer would have had mappings for the same questions of straightforward, complicated, complicated, and straightforward.

So in this experiment respondents had complicated mappings 50 percent of the time. This was so that we could directly compare accuracy

for both kinds of mappings. In ordinary surveys respondents probably experience complicated mappings less than 50 percent of the time, but the actual proportion no doubt varies from respondent to respondent, from question to question, and survey to survey.

Interviewer training. The interviewers were trained in group settings for a total of approximately 90 minutes.[3] We wanted to ensure that standardized and flexible interviewers had the same knowledge about the key concepts in the survey questions, so that any differences in respondent accuracy could not be attributed to one group's greater familiarity with the definitions. So all interviewers first studied the key survey concepts, were quizzed, and discussed the concepts as a group. The interviewers were quizzed on situations that were different from those they would later encounter in the interviews.

Then half the interviewers (11) were selected arbitrarily for separate 1-hour group training in standardized interviewing techniques, and the other half (11) were trained in flexible interviewing techniques. Training consisted of discussion of interviewing theory and role-playing exercises (details are available from the authors), but the interviewers were never informed of the experimental hypotheses.

The standardized interviewers read the relevant sections from the U.S. Department of Commerce's *CPS Interviewing Manual* (1994, pp. A2-6 to A2-8), which are consistent with Fowler and Mangione's (1990) prescriptions. They were trained to read questions exactly as worded and to provide only nondirective probes, but they were never to provide definitions for the survey concepts (we explained that the concepts training had been necessary so that interviewers would be able to judge when respondents had answered a question completely). The probing techniques included rereading the question, providing the response alternatives, and asking the respondents to interpret questions for themselves.

The flexible interviewers were trained to read the questions exactly as worded (just as standardized interviewers do), but then they could say whatever they wanted to assure that the respondent had understood the question as the survey designer had intended. This included reading or paraphrasing all or part of a question, reading or paraphrasing all or part of a definition, and asking questions of the respondent to elicit information so that the interviewer and respondent could jointly reach a correct response. Interviewers could intervene at the respondent's request or voluntarily; that is, interviewers were licensed to intervene whenever they thought the respondent might have misunderstood the question.

3. Ninety minutes of training is far less than the 2–3 days of training that Fowler and Mangione (1990) report is optimal for standardized interviews. But these interviewers were all experienced professionals who had already undergone formal training; our training consisted of additional training in the particulars of an interviewing technique.

All interviewers were instructed to review the concepts before the experimental interviews, and to make sure they had the definitions available during the interviews.

Experimental procedure. When respondents arrived, we obtained their consent to participate and to be audiotaped. An experimenter then read a set of instructions (see appendix B for exact wordings) and answered questions about the procedure. All respondents were told that getting the right answers to the survey questions depended on their paying close attention to details on each page of their scenario packet.

Respondents who would be participating in flexible interviews were given further instructions to encourage active participation. (In our pilot studies, we found that some respondents would only ask questions in the interview when they were explicitly instructed to do so.) They were told to work with the interviewer as a partner to make sure that they had understood the questions in the way the survey designers intended, and that the survey designers' definitions might differ from their own (see appendix B).

Respondents were then left alone to familiarize themselves with the scenarios, which would also be available to them during the telephone interview. When they felt ready, they were called on the telephone by an interviewer and asked the 12 questions. The interviews were unobtrusively audio-recorded.

Results

Implementation of interview techniques. In order to interpret our accuracy results, we need to be sure that interviewers correctly implemented both interviewing techniques. It appears they did. First, standardized interviews contained a high proportion (70 percent) of sequences in which the interviewer asked the question exactly as worded and the respondent immediately provided an answer, followed by no other "moves," as in this example:[4]

> I: Has Dana purchased or had expenses for meats and poultry.
> R: Yes.
> [Interviewer goes on to next question.]

4. In the transcribed excerpts, the following conventions are used: a period between two spaces (.) represents a pause. A colon within a word indicates a lengthened sound. Overlapping speech is enclosed in asterisks. A hyphen at the end of a word ("it-") indicates that the word was cut off. Question marks indicate rising intonation, and utterance-final periods indicate falling or flat intonation, so utterances that have the grammatical form of questions may end with a period. Words or syllables in all capital letters received extra emphasis.

This contrasts with a low proportion (10 percent) of such sequences in the flexible interviews.[5]

Second, far more of what the flexible interviewers said would be considered "illegal" in pure standardized interviewing. Flexible interviewers rephrased all or part of questions, provided all or part of a definition (either verbatim or paraphrased), classified the respondent's description of their circumstances, offered to provide clarification, confirmed or disconfirmed the respondent's interpretation of questions, and requested particular information about the respondent's circumstances.

For example, in the following exchange the flexible interviewer paraphrased the long definition of "household furniture" to answer the respondent's question:

I: Has Kelly purchased or had expenses for household furniture.
R: Um . is a lamp furniture?
I: No sir, we do not include lamps and lighting fixtures.
R: Okay, no.
[Interviewer goes on to next question.]

In pure standardized interviewing, the interviewer should not have answered the respondent's request for clarification, because by doing so she interpreted the survey question for the respondent. In the next example, the flexible interviewer would have violated the rules of standardized interviewing several times:

I: Last week did Pat have more than one job, including part-time, evening or weekend work?
R: Um . s- say that again, because *[laughter]*
I: *La-*
R: She has many clients which she . but it's the same kind of job.
I: Okay. U:h *that would-*
R: *In other* words she is um .
I: Well what kind of work *does she do.*
R: *She ba-* she babysits, and she *has*
I: *O-*
R: different clients.
I: Okay, that would be considered as all one job,
R: *All right*
I: *no matter* how many people she- she worked *for.*
R: *Yes* if it's the same type of job, yes, she has one job *and that's all.*
I: *And is this-* this is the only thing that she does.
R: Yes, and this is it.

5. As a result of technical error, one flexible interview was not audio-recorded and so could not be transcribed. So this analysis, as well as all others involving transcribed interviews, is based on all 21 standardized interviews and 21 of the 22 flexible interviews. In addition, we omitted one response by one respondent from all analyses because he failed to understand the task and answered the question about his own home, rather than the experimental scenario.

I: Okay, so we'll say no for this. She only has one job?
R: She only has one job.
I: And um . [goes on to next question]

This flexible interviewer explicitly tells the respondent how to answer the question given the respondent's description of the scenario. In contrast, a standardized interviewer might have dealt with the respondent's failure to provide a single answer to the question by saying something like, "Let me repeat the question. Last week, did Pat have more than one job, including part-time, evening, and weekend work?"

Such directive interventions occurred for 85 percent of the questions in flexible interviews, but for only 2 percent of the questions in standardized interviews. In fact, all of the 2 percent "illegal" interventions in standardized interviews were incomplete repetitions of the question, but the individual words all appeared in the same order in the full survey question; by some counts these would be legal. In any case, this was the only deviation from standardization that these interviewers ever engaged in. Clearly, the two types of interviews in this experiment were implemented in qualitatively different ways, and much as we had intended.

RESPONSE ACCURACY

Overall response accuracy. Recall that a response in this experiment is accurate if it matches what the official definition dictates. We counted as responses what respondents said (as seen in the transcripts) rather than what the interviewers wrote down, although these almost always coincided. Only three out of the 504 responses were erroneously recorded by interviewers, one in a standardized interview and two in flexible interviews.[6] (For the one interview that was not audio-recorded, the interviewer's tallies could not be verified. In this case, we trusted the interviewer's tallies because they matched the experimenter's tallies during the course of the interview.)

Response accuracy was nearly perfect in both standardized and flexible interviews when the mapping between the question and the scenarios was straightforward, 97 percent and 98 percent, respectively. But the picture was very different when the mapping was complicated. In the standardized interviews, accuracy was very poor, 28 percent. In the flexible interviews, accuracy was nearly 60 percentage points higher, 87 percent. This interaction (mapping $\times$ interview technique) was highly reliable ($F(1, 41)$

6. In the two cases in flexible interviews, the respondent gave a numerical answer (e.g., "three bedrooms") and the interviewer recorded a different number ("one bedroom"). The one recording error in standardized interviews involved the question, "Did Dana have any purchases or expenses for meats and poultry?" In this case, the respondent said "poultry, no meat"; the interviewer coded this as a "no" response, but the respondent probably (correctly) meant "yes," and this is how we scored the response.

= 130.01, $p < .001$). This interaction was also reliable for all questions ($F(1, 11) = 100.74$, $p < .001$).[7] So flexible interviewing led to superior accuracy when it was not obvious to respondents how the questions corresponded to their circumstances, and this was true for all questions.

Respondent characteristics and accuracy. We found no reliable differences in the pattern of results for respondents of different educational levels or races.[8] Of course, our sample was small; this lack of respondent effects would have to be replicated with a larger sample before we can be sure flexible interviewing benefits respondents of various educational levels and races equally.

One characteristic we examined, gender, differentially affected response accuracy (interaction of interview type and gender, $F(1, 39) = 6.20$, $p < .02$). Women were reliably more accurate (97 percent) than men (88 percent) in flexible interviews ($F(1, 20) = 11.74$, $p < .005$), while women and men were equally accurate (61 percent and 64 percent) in standardized interviews ($F(1, 19) = 0.64$, n.s.). Focusing just on flexible interviews, women outperformed men when the mappings were complicated ($F(1, 20) = 6.34$, $p = .02$) and marginally outperformed men when mappings were straightforward ($F(1, 20) = 3.75$, $p < .07$).

There are a number of possible explanations for this gender effect, but our data do not allow us to evaluate any of them. Although this certainly warrants further attention, the gender difference is minor relative to the very large increase in accuracy resulting from the flexible interviewing technique.

Interviewers and accuracy. One goal of standardizing interviews is to reduce interviewer-related variance (Fowler and Mangione 1990). Because flexible interviews involve extensive probing, one might expect greater interviewer effects for flexible interviews than for standardized interviews. One reason to expect this is because Mangione, Fowler, and Louis (1992) found that the questions in standardized interviews requiring the most probing were most likely to create interviewer effects.

This turned out not to be the case here. Interviewer variance was no greater in flexible interviews than in standardized interviews (interaction of interviewers and interview type, $F(20, 42) = 1.11$, n.s.). Of course, there were only two respondents per interviewer, and so we cannot compute a measure of interviewer-related variance like rho (Fowler and Mangione 1990), which requires that the respondents assigned to interviewers represent the sample as a whole. Our data do not rule out the possibility that in a larger sample flexible interviewing could lead to greater interviewer effects than standardized interviewing.

7. This was computed by treating questions (rather than respondents) as the random factor in an analysis of variance.

8. We excluded three respondents from the analysis of race effects because their racial grouping (Asian) was only represented in the flexible interviewing condition.

NATURE OF THE INTERACTION

Accuracy of flexible interviewers' interventions. One of the potential dangers of flexible interviewing is that interviewers can mislead respondents. That is, even if interviewers sometimes provide information that helps respondents to produce accurate answers, interviewers may also provide information that can lead respondents astray.

To test this, we identified those cases in which flexible interviewers provided explicit, directive information (49 percent of the cases where the mappings were straightforward, and 75 percent of the cases where the mappings were complicated). We used a stringent criterion for accuracy: we considered any information that did not appear in the definitions to be inaccurate. Take this example:

> I: How many hours per week does Mindy usually work at her job.
> R: She usually works fifty- a average of fifty hours a week.
> I: Fifty hours a week?
> R: Mm-hm?
> I: That the average work week?
> R: Yeah for the last six months thos- that's what's- that's . was her average work hour week. Fifty hours.
> I: All right [continues]

In this case we counted the interviewer's comment "That the average work week?" as inaccurate, because it endorses the incorrect (for this survey) interpretation of "average" for "usually," rather than the correct interpretation of "most frequent" (see appendix A).

We also counted as inaccurate definitions that interviewers improvised when no official definition had been provided, as in this example, where the interviewer invents a definition for "farm":

> R: . What do you mean by a farm?
> I: A farm? It would be a . you know any . farm that would be producing . any- any, yeah anything uh could be cattle or uh . vegeta- vegetables, or orchard . uh that uh would be producing for . income . for the household it wouldn't be a . a farm just for the uh . household use only.

Under this stringent criterion, flexible interviewers provided only accurate help in 93 percent of the cases where they provided any help. When interviewers provided accurate information, respondents answered accurately 87 percent of the time and inaccurately the remaining 6 percent of the time. On the 7 percent of occasions when interviewers provided any inaccurate information, respondents still produced the correct answer 4 percent of the time and produced incorrect answers 3 percent of the time. So flexible interviewers generally provided highly accurate information, and providing inaccurate information did not necessarily lead respondents to produce incorrect answers.

All 11 flexible interviewers presented far more accurate than inaccurate

information. Four provided perfectly accurate information. Four others presented inaccurate information on one question out of the 24 total questions they asked—a rate of 4.2 percent inaccuracy. Two interviewers presented inaccurate information on two questions (a rate of 8.3 percent), and the remaining interviewer on three questions (a rate of 12.5 percent).

When did flexible interviewers intervene? Interviewers sometimes intervened because respondents asked for help, and sometimes they intervened voluntarily. They intervened for both straightforward and complicated mappings, but more often for complicated (88 percent) than for straightforward mappings (51 percent). This partially reflects the fact that respondents almost never asked for help with straightforward mappings.

Recall that the interviewers had no evidence at the outset whether respondents were faced with complicated or straightforward mappings. So why did they intervene voluntarily? Sometimes it was because respondents had displayed uncertainty in their answers or failed to answer the questions definitively. The most common way that respondents showed they were uncertain was to describe their situation; they did this for 58 percent (37 of 64) of the complicated cases where interviewers intervened voluntarily. In this example, the interviewer begins to provide substantive help after the respondent describes the scenario at length:

I: How many hours per week does she u- does Mindy usually work at her job.
R: Well Mindy's job schedule varies.
I: *mm-hm*
R: *as far as* what she usually works,
I: Mm-hm
R: in the average of how many hours she works there's a difference, um . in the past six months, three of the months she worked fifty hours, but then there are two that was you know one at forty and one at ten. So her average would be about forty hours a week, but I would say she usually works . between forty . forty and fifty, there's an odd week here and there, but USUALLY, between forty and fifty.
I: Okay you mentioned of the last six months for . half of them she worked fifty hours.
R: Yes.
I: And . fifty percent-
R: Actually there are only five months here, *yes.*
I: *Five* months?
R: So yes, over- over three fifths of the time she works fifty hours a week.
I: Okay. So fifty percent of the time or more
R: She's working fifty hours a week.
I: Okay and we would consider fifty percent of the time or more or the most frequent schedule during the por- past four or five month to be her usual number of hours and you said that was fifty hours for *most* of the time.
R: *Yes.* Yes.
I: Okay.

Interviewers also intervened voluntarily when respondents asked them to repeat the question (9 percent of cases, 6 of 64), when respondents explicitly said they were unsure about the answer (6 percent of cases, 4 of 64), and in a variety of less frequent circumstances for the remaining 27 percent (17 of 64) of the cases.

A closer inspection of the interaction shows that interviewers really did increase response accuracy by providing unsolicited help. Flexible respondents only requested help for 38 percent of the complicated mappings, but, as we have seen, they were accurate for 87 percent of them. They were virtually as accurate when the interviewers volunteered help (86 percent accuracy, 55 out of 64 cases) as when they explicitly requested help (94 percent accuracy, 46 out of 49 cases). In contrast, when interviewers failed to provide any help at all (11 complicated-mapping cases), respondents only produced 4 accurate answers, a rate of 34 percent. This is nearly as poor as the standardized respondents' 28 percent accuracy rate for complicated mappings.

While flexible respondents got help when they needed it, they also got help when, it would seem, they did not. Respondents almost never asked for help for straightforward mappings (1 percent of the time), but they nonetheless received unsolicited help for 51 percent of the straightforward cases. One cost of flexible interviewing may be that interviewers provide a substantial amount of unnecessary help in addition to the needed help.

Were respondents more accurate if they explicitly asked for help? We see no evidence that they were. Respondents who asked for help frequently were no more accurate than respondents who asked for help rarely ($r = .15$, n.s.). In a sense, it did not matter how much respondents explicitly asked for help, because interviewers provided help whether or not respondents asked for it. In other words, respondents in flexible interviews all benefited from flexibility, but only sometimes as a result of their own initiative.

Effectiveness of standardized (nondirective) interventions. According to proponents of standardization, well-trained standardized interviewers should be able to get respondents to answer appropriately by using neutral probing techniques (Fowler and Mangione 1990). In the vast majority of our standardized cases (70 percent), interviewers did not probe at all; respondents answered the questions directly. Respondents almost never explicitly asked for help interpreting the questions (only on four questions out of all 252 asked), and when they did, they were told that the interpretation was up to them.

But respondents did provide less explicit evidence of uncertainty. In these cases, interviewers sometimes provided effective and legal standardized probes that clearly led to improved response accuracy, as in this next example from a question with a complicated mapping. The interviewer

leaves the interpretation up to the respondent and repeats the question; the respondent then produces the correct answer:

I: How many people live in this house.
R: Currently? Or
I: Okay uh we need your interpretation.
R: *Um*
I: *How many* people . live in this house.
R: Three.
I: Three.
I: Okay, [continues]

One probing technique that some interviewers used was to elicit a correct response by repeating the question with different contrastive stress, as in these two examples from different interviewers and respondents:

I: How many bedrooms are there in THIS house.
R: Uh, there are two bedrooms. And one den is being used as a bedroom.
I: How many BEDROOMS are there in this house.
R: Two.
I: [continues]

I: How many hours per week does Mindy usually work at her job.
R: Um: it varies, but she seems to average . m: . about thirty eight hours.
I: How many hours per week does she USUALLY . work at her job.
R: Um, fifty hours.
I: [continues]

But this strategy did not always lead to a correct response, as in this case:

I: *How* many- how many hours per week does Mindy usually work at her job?
R: Um during which month. In general?
I: Uh, this would be uh how many hours does she USually work at her job.
R: Okay,
I: *Cou-*
R: *U:*m
I: Could you tell me
R: Uh, thirty hours.
I: [continues]

So neutral probing was not always effective, as we also saw in the response accuracy data.

Although our standardized interviewers overwhelmingly used only legal probing techniques, it seems to us on closer examination that at least some of these probes are not, strictly speaking, neutral: they convey information about the official definitions. Consider this example:

I: Has Alexander purchased or had expenses for college tuition or fixed fees.
R: Um he's go:t . um tuition for secretarial school.

I: Pardon me, I- I didn't .
R: He has tuition for secretarial school.
I: Has Alexander purchased or had expenses for COLLEGE tuition
R: No.
I: or fixed fees.
R: No.
I: [continues]

When the interviewer repeated the question with contrastive stress, might this not have signaled to the respondent that, for current purposes, secretarial schools should not be considered colleges? Contrastive stress is one technique speakers in ordinary conversation use to call attention to novel or unexpected features of their utterances (see Chafe 1976). In this example, the respondent produced the correct response, but we have no guarantee that the respondent would have made the same choice without the unscripted emphasis used by the interviewer in this probe.

In fact, virtually all the legal moves a standardized interviewer can make are, strictly speaking, not neutral, because they can convey information to the respondent about how he should interpret the question. When an interviewer repeats a question after a respondent has given a tentative answer, she may be signaling to the respondent that his answer was wrong. Even the choice not to probe implies the interviewer's willingness to accept the respondent's interpretation of the question; if the interviewer moves on without probing, she has implicitly signaled that the respondent's interpretation is indeed the appropriate one (for further discussion, see Clark and Schaefer 1989; Clark and Schober 1991; Schober 1998b; Schober and Conrad 1998; Schwarz 1994, 1996, among others).

DURATION

Duration of interviews. Although flexible interviewing led to massive improvements in accuracy, and although flexible interviewers rarely misled the respondents, the technique did have a significant cost: flexible interviews took much longer than standardized interviews did. The median time to complete flexible interviews was 11.47 minutes, compared to 3.41 minutes for standardized interviews; one flexible interview lasted over 35 minutes, and the shortest flexible interview took as long as the longest standardized interview (about 6 minutes). As shown by the total number of words per question uttered by respondents and interviewers, flexible interviews took longer than standardized interviews regardless of whether mappings were complicated (933 vs. 211 words) or straightforward (727 vs. 158 words) (interaction of interview type and mapping, $F(1, 40) = 2.56$, n.s.).

So there is a clear trade-off between improved accuracy and saving time. But this trade-off may be less extreme than it seems. First, our inter-

viewers were new to the definitions and the technique. As a result, some flexible interviewers were not adept at focusing on just the relevant parts of definitions and read entire lengthy definitions verbatim, as in this inelegant example:

I: Last week, did Pat have more than one job, including part-time, evening or weekend work?
R: W:hat . is a job.
I: All right. Uh, a job exists when there is a definite arrangement for regular work every week, or every month, for pay or other compensation. By other compensation that would be profits, anticipated profits, or pay in kind, such as room and board. A formal, defin- definite arrangement with one or more employers to work on a continuing basis for a specified number of hours per week or days per month, but on an irregular schedule, during the week or month is also a job. It is possible for individuals to have more than one employer, but only one job. If an individual does the same type of work for more than one employer in an occupation where it is common to have more than one employer, do not consider the individual a multiple jobholder. Examples improve . in- include rather, private households or domestic workers . including babysitters, chauffeurs, gardeners, handypersons, cooks, and maids.
R: You said do NOT include? . babysitters?
I: Uh, let's see: .
R: The last sentence.
I: All right . let's see do not . If an individual does the same type of work for more than one employer, in an occupation where it is common to have more than one employer, do not consider the individual a multiple jobholder. Examples include private household or domestic workers . include- including babysitters, chauffeurs, gardeners, handypersons, cooks, and maids.
R: Okay. And the question again was .
I: Last *week*
R: *just*
I: did Pat have more than one job, including part-time, evening, or weekend work.
R: N:o, she had one job.
I: [goes on to next question]

In comparison, the interviewer in this example gets right to the point:

I: Last week, did Pat have more than one job, including part-time, evening or weekend work.
R: Last week?
I: Yes.
R: U:h more than one job?
I: Yes.
R: She had many jobs.
I: What does she do.
R: What does P- she babysits.
I: Okay, so she works for different employers.
R: Yes, different families.
I: Okay, she doesn't do this in her home.

R: No?
I: Okay. Did she do anything else besides her babysitting job.
R: No?
I: Okay. [goes on to next question]

Here the interviewer asks the respondent to describe Pat's circumstances as soon as the respondent says, "She had many jobs"; this might have signaled that the respondent was unsure about the interpretation of the question (see, e.g., Brennan and Clark [1996] and Schober [1998a] on the implications of using different wording than one's conversational partner—in this case, "many" vs. "more than one"). With one phrase, the interviewer elegantly classifies Pat's work as one job by calling it "her babysitting job." More experienced flexible interviewers might use strategies like these, which in turn might lead to shorter flexible interviews.

Second, our experiment implemented pure versions of the interviewing techniques. According to common wisdom, ordinary interviews deviate from standardization and thus take longer than pure standardized interviews. To evaluate the potential impact of moving to flexible interviewing, the appropriate comparison is between our flexible interviews and ordinary interviews, rather than our flexible and standardized interviews. The increase in duration for flexible interviews may therefore be smaller in practice than in our study.

Discussion

As Tourangeau (1990, p. 251) puts it, "it is not a foregone conclusion that the costs of standardization outweigh the gains or that the gains can be preserved while the costs are reduced." As a first attempt to address the issue, our study shows that there are circumstances under which the costs of standardization do outweigh the gains, if high response accuracy is the goal.

Flexible interviewing led to nearly 60 percent greater accuracy when the mapping between the question and the respondent's situation was complicated. This large accuracy increase was obtained without lengthy interviewer training, using typical telephone interviewers, and without increasing interviewer effects over what we found in the standardized interviews. Flexible interviewing led to greater response accuracy for respondents of different races, genders, and levels of educational attainment.

But this accuracy came at a real cost—a more than threefold increase in duration. As we have noted, however, our flexible interviews may be longer than they would be in actual practice, and our standardized interviews may be shorter than ordinary "standardized" interviews; this requires further investigation in real survey settings.

We believe flexible interviewing is a promising alternative to explore.

But we agree with Schaeffer's (1991, p. 368) point: "Reforms that ignore the justification for standardization run the risk of repeating old mistakes." We do not advocate implementing flexible interviewing without further careful research with different samples and different kinds of surveys.

In particular, a number of questions need to be addressed. How far do our results extend beyond the laboratory, when the frequency of complicated mappings is not controlled? Can flexible interviewing work for all interviewers and all respondents? Do different versions of flexible interviewing affect response accuracy differently and have different costs? How would flexible interviewing affect response accuracy for attitude or opinion questions?

In any case, the results of our study suggest at least the following conclusions.

1. *Mappings are a potential source of measurement error for any question.* Every survey question contains terms that have the potential to be understood differently than the survey designers intended, even if the questions have been pretested. This is because respondents' circumstances may not map onto the official definitions in a straightforward way. Mapping problems differ from the question-meaning (word and sentence) problems that pretesting can effectively address, because they involve the *correspondence* between official question meaning and respondents' personal circumstances, and this is hard (if not impossible) to anticipate.

Our position is that while pretesting and wording changes are necessary to reduce predictable misunderstandings, they cannot accommodate all complicated mappings. Respondents' circumstances are too varied; official definitions for the words in a question can be too long and complex, and they will never match every respondent's intuitions about what words mean. Flexible interviewing may turn out to be a solution when complicated mappings are frequent (provided that official definitions have been developed); when complicated mappings are rare, flexible interviewing may not be worth the expense.

2. *Different interviewing techniques may be appropriate for different circumstances.* We have shown that standardized interviewing techniques can lead to measurable inaccuracy in responding, and that, under some circumstances, flexible interviewing can lead to measurably improved response accuracy. But flexible interviewing is no panacea—it has real costs. Our data begin to quantify the trade-offs that survey researchers face as they weigh their simultaneous needs for accurate responses, speedy interviews, reasonable interviewer training costs, and reasonable question-development costs (including developing definitions of question concepts), among others. In some circumstances, cost constraints may require survey researchers to accept the reduced certainty of response accuracy inherent in standardized interviewing. In other circumstances, accurate responses may be worth any price.

Appendix A

Questions and Definitions of Key Concepts[9]

HOUSING QUESTIONS (FROM CPI HOUSING SURVEY)

1. How many *bedrooms* are there in this house?

A bedroom is a finished room specifically designed by the owner to be used for sleeping. A bedroom does NOT have to be used for sleeping in order to qualify as a bedroom. For example, a bedroom that is being used as an office should be counted as a bedroom.

Do NOT count as a bedroom any room that was designed for another purpose but is being used as a bedroom. For example, a den being used as a bedroom is still a den and should not be counted as a bedroom.

Do NOT count as a bedroom any dens, living rooms, or other rooms that can be converted at night for sleeping.

Do NOT count any bedroom that the renter is denied access to or use of by the owner.

A one-room efficiency apartment does not have a bedroom.

2. This question has two parts. How many *full bathrooms* are there in this house? How many *half bathrooms* are there?

A full bathroom has (1) a flush toilet, (2) a bathtub or shower, and (3) a sink or washbasin with running water. Bathrooms that contain all of the above items, whether separated by a partition or door, are to be considered a full bathroom.

A half bathroom has any two of these three items: (1) a flush toilet, (2) a bathtub or shower, and (3) a sink or washbasin with running water.

If the only bathroom facilities do not meet the definition of a full or half bath, code zero. (For example, if there is only a flush toilet in a room.)

If a bathroom is shared by the occupants of more than one housing unit, the bathroom is included with the unit from which it is most easily reached.

3. How many *other rooms* are there, other than bedrooms and bathrooms?

Include whole rooms such as living rooms, dining rooms, kitchens, lodger's rooms, finished basements or attic rooms, recreation rooms, and permanently enclosed sun porches. Rooms used for offices by a person living in the unit are also included in this survey. Rooms are counted even if they are not used.

Do NOT include bedrooms, bathrooms, unfinished attics or basements, halls, foyers or vestibules, balconies, closets, alcoves, pantries, strip or pullman kitchens, laundry or furnace rooms, open porches, and unfinished spaces used for storage.

A partially divided room, such as a dinette next to a kitchen or living room, is a separate room ONLY if there is a PERMANENT PARTITION FROM FLOOR TO CEILING BETWEEN THE TWO AREAS. An L-shaped room, a

9. Key concepts are italicized.

''great'' room, or a step-down is therefore counted as one room unless there is a permanent partition dividing the room into parts.

If a room is used by occupants of more than one unit, the room is included with the unit from which it is most easily reached.

Do NOT count any rooms that the renter is denied access to or use of by the owner. Do count rooms REGARDLESS of their year-round usability.

Bathrooms: exclude all bathrooms. While some rooms, such as a small room with only a wash basin, do not meet the definition of a bathroom, they are also to be excluded from the count of other rooms.

4. How many people *live in this house?*

A person is considered to be living in a housing unit even if the person is not present at the time of the survey. Live-in servants or other employees, lodgers, and members of the household temporarily away from the unit on business or vacation are included in the count.

Do NOT count any people who would normally consider this their (legal) address but who are LIVING away on business, in the armed forces, or attending school (such as boarding school or college).

Do NOT count overnight lodgers, guests and visitors. Do NOT count day employees who live elsewhere.

WORK QUESTIONS (FROM CPS SURVEY)

1. Does anyone in this household have a *business* or a farm?

A business exists when one or more of the following conditions is met: Machinery or equipment of substantial value is used in conducting the business, or an office, store, or other place of business is maintained, or the business is advertised by: listing in the classified section of the telephone book, or displaying a sign, or distributing cards or leaflets or otherwise publicizing that the work or service is offered to the general public.

2. Last week, did Chris do any work for *pay?*

Include piece rate income as earnings. Persons working in garment making or food packaging often receive this type of income. Also count college assistantships and fellowships and on the job training as earnings.

DO NOT INCLUDE PAY IN KIND, such as food or lodging for work, or expense accounts as earnings.

3. Last week, did Pat have *more than one job,* including part-time, evening or weekend work?

A job exists when there is a definite arrangement for regular work every week, or every month, for pay or other compensation (e.g., profits, anticipated profits, or pay in kind, such as room and board). A formal, definite arrangement with one or more employers to work on a continuing basis for a specified number of hours per week or days per month, but on an irregular schedule during the week or month, is also a job.

It is possible for individuals to have more than one employer, but only one job. If an individual does the same type of work for more than one employer in an occupation where it is common to have more than one employer, do not consider the individual a multiple jobholder. Examples include private household or domestic workers including babysitters, chauffeurs, gardeners, handypersons, cooks, and maids.

4. How many hours per week does Mindy *usually* work at her job?

50 percent of the time or more, or the most frequent schedule during the past 4 or 5 months.

PURCHASE QUESTIONS (FROM CPOPS SURVEY)

1. Has Carla purchased or had expenses for *car tires?*

New, recapped, or retreaded tires for automobiles. Do not include tires for vans and trucks.

2. Has Alexander purchased or had expenses for *college tuition or fixed fees?*

Tuition and fixed fees paid to public or private institutions offering credit beyond the high school level. Do not include payments to vocationally oriented schools such as business, technical, trade, or secretarial; do not include payments for room and board, books, lab fees, etc.

3. Has Kelly purchased or had expenses for *household furniture?*

Tables, chairs, footstools, sofas, china cabinets, utility carts, bars, room dividers, bookcases, desks, beds, mattresses, box springs, chests of drawers, night tables, wardrobes, and unfinished furniture. Do not include TV, radio, and other sound equipment, lamps and lighting fixtures, outdoor furniture, infants' furniture, or appliances.

4. Has Dana purchased or had expenses for *meats and poultry?*

Beef, lamb, pork, game; organ meats, such as kidneys, sweetbreads, chitterlings, heart, tongue; sausages and luncheon meats; poultry, such as chicken, turkey, pheasant, goose, duck. Include canned ham. Do not include other canned meats and canned poultry, or any prepared meats and poultry.

Appendix B

Instructions to Respondents

INSTRUCTIONS

We'll be asking you to answer 12 survey questions, each about a different fictional situation. You'll be using a 12-page packet, where each page describes one of

these situations. For the first question, you should use the information on the first page. For the second question, use the information on the second page, and so on. Sometimes a page contains a very short story about someone's living or working situation; sometimes you'll see a floor plan of a house or apartment; sometimes you'll see a receipt from a purchase.

You'll be talking on the phone with a professional survey interviewer. The interviewer will know that you are answering these questions based on what is in your packet, but they won't know what is on each page of your packet. That is, they will *not* be seeing a packet that looks like yours, but only a list of questions, and so they don't know what the right answer is. We would like you to answer using the information available on each page.

Before the interview starts, we would like you to get to know the situations by reading each page *very* carefully. You don't need to memorize the information on each page; during the interview you should use the packet to help you answer the questions.

According to the definitions of this agency (the Bureau of Labor Statistics), there *is* a correct answer for each question. Sometimes getting the right answer depends on your having paid attention to details on each page.

It is VERY important that you fully understand these instructions. If you have ANY questions, please ask them now.

Now, please turn to your packet and begin familiarizing yourself.

ADDITIONAL INSTRUCTIONS FOR FLEXIBLE INTERVIEWS

Sometimes these survey questions use ordinary words with slightly different meanings than you may be used to. This is because surveys sometimes need to have technical definitions different from ordinary definitions. You shouldn't feel at all reluctant to ask if you aren't sure what we mean by a perfectly ordinary word. In fact, we WANT you to ask if you have ANY uncertainty about how to interpret the question—even if this feels silly to you. So, even if you know perfectly well what a "person" or a "house" is, if a question includes those words, you may need to ask the interviewer for a definition. The interviewer will be more than happy to help you as much as possible.

It may be that if you don't ask about word meanings, you won't be able to get the right answer, because you may be thinking about the question differently than the people who wrote it. For example, imagine that you see a shopping receipt that shows that Gina bought butter. If the interviewer asked you, "Did Gina buy any fats or oils?" you might want to say yes, because butter seems to be a fat. But the official definition of "fats or oils" excludes butter, and so the correct answer would be no. If you didn't ask whether butter is a fat or not, you probably would get the wrong answer.

It is VERY important that you fully understand these instructions. If you have ANY questions, please ask them now.

Now, please turn to your packet and begin familiarizing yourself with the situations.

References

Beatty, Paul. 1995. "Understanding the Standardized/Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11:147–60.

Belson, William A. 1981. *The Design and Understanding of Survey Questions.* Aldershot: Gower.

———. 1986. *Validity in Survey Research.* Aldershot: Gower.

Brennan, Susan E. 1990. "Seeking and Providing Evidence for Mutual Understanding." Doctoral dissertation, Stanford University.

Brennan, Susan E., and Herbert H. Clark. 1996. "Conceptual Pacts and Lexical Choice in Conversation." *Journal of Experimental Psychology: Learning, Memory and Cognition* 22(6):1482–93.

Briggs, Charles L. 1986. *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research.* Cambridge: Cambridge University Press.

Chafe, Wallace L. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View." In *Subject and Topic,* ed. Charles N. Li, pp. 25–56. New York: Academic Press.

Cicourel, Aaron V. 1973. *Cognitive Sociology: Language and Meaning in Social Interaction.* New York: Free Press.

Clark, Herbert H. 1992. *Arenas of Language Use.* Chicago: University of Chicago Press.

———. 1996. *Using Language.* Cambridge: Cambridge University Press.

Clark, Herbert H., and Edward F. Schaefer. 1989. "Contributing to Discourse." *Cognitive Science* 13:259–94.

Clark, Herbert H., and Michael F. Schober. 1991. "Asking Questions and Influencing Answers." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys,* ed. Judith M. Tanur, pp. 15–48. New York: Russell Sage Foundation.

Clark, Herbert H., and Deanna Wilkes-Gibbs. 1986. "Referring as a Collaborative Process." *Cognition* 22:1–39.

Fowler, Floyd J. 1991. "Reducing Interviewer-Related Error through Interviewer Training, Supervision, and Other Means." In *Measurement Errors in Surveys,* ed. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, pp. 259–78. New York: Wiley.

Fowler, Floyd J., and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error.* Newbury Park, CA: Sage.

Goffman, Erving. 1981. *Forms of Talk.* Philadelphia: University of Pennsylvania Press.

Goodwin, Charles. 1981. *Conversational Organization: Interaction between Speakers and Hearers.* New York: Academic Press.

Gumperz, John J. 1982. *Discourse Strategies.* Cambridge: Cambridge University Press.

Holstein, James A., and Jaber F. Gubrium. 1995. *The Active Interview.* Thousand Oaks, CA: Sage.

Hyman, Herbert H. 1954. *Interviewing in Social Research.* Chicago: University of Chicago Press.

Kovar, Mary G., and Patricia Royston. 1990. "Comment on Suchman and Jordan." *Journal of the American Statistical Association* 85(409):246–47.

Krauss, Robert M., and Susan R. Fussell. 1996. "Social Psychological Models of Interpersonal Communication." In *Social Psychology: Handbook of Basic Principles,* ed. E. Tory Higgins and Arie Kruglanski, pp. 655–701. New York: Guilford.

Kvale, Steinar. 1994. "Ten Standard Objections to Qualitative Research Interviews." *Journal of Phenomenological Psychology* 25:147–73.

Mangione, Thomas W., Floyd J. Fowler, and T. A. Louis. 1992. "Question Characteristics and Interviewer Effects." *Journal of Official Statistics* 8:293–307.

Mishler, Elliot G. 1986. *Research Interviewing.* Cambridge, MA: Harvard University Press.

Rogoff, Barbara. 1990. *Apprenticeship in Thinking: Cognitive Development in Social Context.* New York: Oxford University Press.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking in Conversation." *Language* 50: 696–735.

Schaeffer, Nora Cate. 1991. "Conversation with a Purpose—or Conversation? Interaction in the Standardized Interview." In *Measurement Errors in Surveys,* ed. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, pp. 367–91. New York: Wiley.

Schegloff, Emanuel A. 1984. "On Some Questions and Ambiguities in Conversation." In *Structures of Social Action,* ed. J. Maxwell Atkinson and John Heritage, pp. 28–52. Cambridge: Cambridge University Press.

Schiffrin, Deborah. 1994. *Approaches to Discourse.* Cambridge, MA: Basil Blackwell.

Schober, Michael F. 1998a. "Different Kinds of Conversational Perspective-Taking." In *Social and Cognitive Psychological Approaches to Interpersonal Communication,* ed. Susan R. Fussell and Roger J. Kreuz. Mahwah, NJ: Erlbaum.

———. 1998b. "Making Sense of Questions: An Interactional Approach." In *Cognition and Survey Research,* ed. Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau. New York: Wiley.

Schober, Michael F., and Herbert H. Clark. 1989. "Understanding by Addressees and Overhearers." *Cognitive Psychology* 21:211–32.

Schober, Michael F., and Frederick G. Conrad. 1998. "A Collaborative View of Standardized Survey Interviews." To appear in *Interaction in the Standardized Survey Interview,* ed. Hanneke Houtkoop-Steenstra, Douglas Maynard, Nora Cate Schaeffer, and Johannes van der Zouwen.

Schwarz, Norbert. 1994. "Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation." In *Advances in Experimental Social Psychology,* vol. 26, ed. M. Zanna, pp. 123–62. San Diego, CA: Academic Press.

———. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation.* Mahwah, NJ: Erlbaum.

Suchman, Lucy, and Brigitte Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85(409):232–53.

———. 1991. "Validity and the Collaborative Construction of Meaning in Face-to-Face Surveys." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys,* ed. Judith M. Tanur, pp. 241–67. New York: Russell Sage Foundation.

Tannen, Deborah. 1989. *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse.* Cambridge: Cambridge University Press.

Tourangeau, Roger. 1990. "Comment on Suchman and Jordan." *Journal of the American Statistical Association* 85(409):250–51.

U.S. Department of Commerce. 1994. *Current Population Survey Interviewing Manual* (CPS-250). Washington, DC: Bureau of the Census.

van der Zouwen, Johannes, Wil Dijkstra, and Johannes H. Smit. 1991. "Studying Respondent-Interviewer Interaction: The Relationship between Interviewing Style, Interviewer Behavior, and Response Behavior." In *Measurement Errors in Surveys,* ed. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, pp. 419–37. New York: Wiley.

Wentland, Ellen J. 1993. *Survey Responses: An Evaluation of Their Validity.* San Diego, CA: Academic Press.