

CONVERSATIONAL INTERVIEWING AND DATA QUALITY

Frederick G. Conrad
Bureau of Labor Statistics

Michael F. Schober,
New School for Social Research

Abstract

Standardized interviewing is widely practiced because it promises to reduce interviewer related error and because it is cheap. Yet the technique cannot guarantee uniform understanding of questions and, thus, may reduce data comparability. Conversational interviewing may standardize the meaning of questions by allowing interviewers to clarify survey concepts, but it cannot guarantee uniform interviewer behavior. In a series of studies, we have found that conversational interviewing improves response accuracy when respondents' circumstances are atypical (e.g. Should a lamp purchase be counted as a furniture purchase?) but requires additional time to clarify concepts. In addition, respondents are not always willing to seek clarification when they need it. Survey researchers have the option of keeping interviews short and risking some misunderstanding or investing more time to be sure questions are understood as intended.

The Standardization Debate

Standardized survey interviewing is widely practiced in all sectors of the survey industry. It has evolved to reduce interviewer related error. The goal is to provide all respondents with the exact same wording and without commentary that could bias answers (Fowler and Mangione, 1990). If questions have been pretested to correct major problems (Kovar and Royston, 1990) and they are presented identically from one occasion to the next, then most respondents should interpret them the same way. In addition to its promise of reducing interviewer related error it is a relatively fast and cheap way to collect survey data (Beatty, 1995). Yet, standardization has been criticized on the grounds that it may prevent interviewers from resolving certain respondent misunderstandings and, as a result, actually increase measurement error (Suchman and Jordan, 1990). This is because the same question may be interpreted differently by different respondents, thus, reducing data comparability.

Conversational interviewing is an alternative approach, designed to assure that all respondents understand questions as intended. Interviewers say what is needed to help respondents correctly interpret the questions. It is not wording, therefore, but meaning that is standardized in a conversational interview (Suchman and Jordan, 1990). The approach is based on a view of communication that requires partners to *collaborate*, to converse about what is being said until they are confident they adequately understand each other (see, e.g. Cicourel, 1973; Clark, 1992; Schegloff, 1984; Tannen, 1989). In conversational interviews, it is essential that the interviewer and respondent talk about the meaning of the questions because the questions might otherwise mean one thing to the

survey sponsors and something else to the respondent. Conversational interviewing has been criticized on the grounds that it makes it possible for interviewers to lead (and mislead) respondents (Fowler and Mangione, 1990; Tourangeau, 1990, p. 251).

Evaluating the Costs and Benefits of Standardized and Conversational Interviewing

We have carried out a series of experimental studies, both in the laboratory (Schober and Conrad, 1997; Schober and Conrad, 1998; Schober, Conrad and Fricker, 1999; Conrad and Schober, 1998) and the field (Conrad and Schober, 1999) to quantify the relative advantages of these two interviewing approaches and some of their variants. In all cases we have assessed the impact of interviewing techniques on response accuracy. In the laboratory studies that we discuss below, we directly assessed response accuracy (the extent to which answers fit with survey concepts) by using fictional scenarios. Respondents answered on the basis of purchase receipts, textual vignettes, and floor plans that were not available to the interviewers. In the field study, we used two indirect measures of accuracy: response change between interviews and respondents' explanations for their answers.

Typical and atypical situations. It is possible for someone to ask you a question which is hard to understand, even though you are familiar with all of the words. This can happen because you're not sure how the words correspond to your particular circumstances. Suppose you are asked the following question from the Point of Purchase Survey conducted by the Bureau of Labor Statistics (BLS). "During the last year, have you purchased or had expenses for household furniture?" Suppose, further, that you purchased a lamp during this period. Is a lamp household furniture? It is less typical of furniture than are chairs but it *could* be considered furniture – just atypical furniture. If you had purchased a chair, it would be a straightforward matter to answer "yes." In the case of a lamp, it is more complicated; it all depends on how BLS defines "household furniture" for this question¹.

The typicality of the situations about which respondents answer is important for several reasons. The type of interview could make a substantial difference when the situations are atypical. If a respondent can easily determine whether the survey sponsors consider lamps to be furniture, she should be able to respond accurately but if it is difficult or impossible for her to do this, her accuracy may suffer. If a respondent in a conversational interview asks whether or not she should include lamps, the interviewer could tell the respondent that, for the purposes of this survey, lamps are not considered to be furniture. If a respondent in a strictly standardized interview asked the same question the

¹ When we have written about this work elsewhere we have used the terms "straightforward mappings" and "complicated mappings." These correspond to what we are calling typical and atypical situations here.

interviewer would be obliged² to answer with a “neutral” probe such as “Let me repeat the question” or “We want *your* interpretation.”³

Alternatively, it is possible that people’s circumstances are usually typical of the survey concepts about which they have been asked – for example, they have bought more chairs than lamps – and so there is not an overwhelming need for them to talk about the meaning of the concepts with interviewers. In this case, standardized interviews should regularly produce accurate responses. In our lab studies, half of the scenarios on which respondents based their answers described typical situations and half described atypical situations. A purchase receipt for an end table described a typical situation while an otherwise identical receipt for a floor lamp described an atypical situation. However, in our field study, respondents answered about their own lives so the typicality of their circumstances was not under our control. Any advantage for conversational over standardized interviewing in this study would indicate that respondents’ circumstances are atypical often enough that the type of interviewing could matter.

Questions and concepts. In most of our studies we have asked 10 or 12 questions, all from ongoing federal surveys. In all cases the sponsoring agencies publish definitions for the key concepts. Examples of these concepts include “business” and “more than one job” from the Current Population Survey (CPS), “inside home maintenance and repair” and “household furniture” from the Point of Purchase Survey (POPS), and “other rooms, beside bedrooms and bathrooms” and “live in your home” from the Consumer Price Index – Housing Survey (CPI-Housing). These definitions are long: average number of words for CPS concepts is 47, for POPS 31 and for the CPI-Housing Survey 103.

When and how to provide clarification. In the strictly standardized interviews that we have administered, there is no way for interviewers to directly clarify the survey concepts. They can use neutral probes to signal some aspects of meaning but otherwise definitions – if they are to be conveyed – must be included in the question or in a series of questions. This is often impractical for the lengthy definitions that are typical of federal surveys. In our first laboratory study (Schober and Conrad, 1997) we contrasted this strict version of standardization with conversational interviews in which, after reading the questions as worded, interviewers could define concepts whenever they judged that it might help respondents. This could involve defining concepts when respondents specifically asked for clarification or when the interviewers thought the respondents might be confused or have misunderstood the question. Interviewers could provide all of a definition or just the relevant part(s); they could present the definitions verbatim or they could paraphrase them. They could say whatever was necessary to help respondents understand the question as it was intended.

² We recognize that many survey organizations which subscribe to the philosophy of standardization allow interviewers to provide this type of clarification. However, this practice departs from pure standardization in that some respondents may obtain the information and some may not (Fowler and Mangione, 1990, p. 21). Our research tests the pros and cons of the *theory* of standardization.

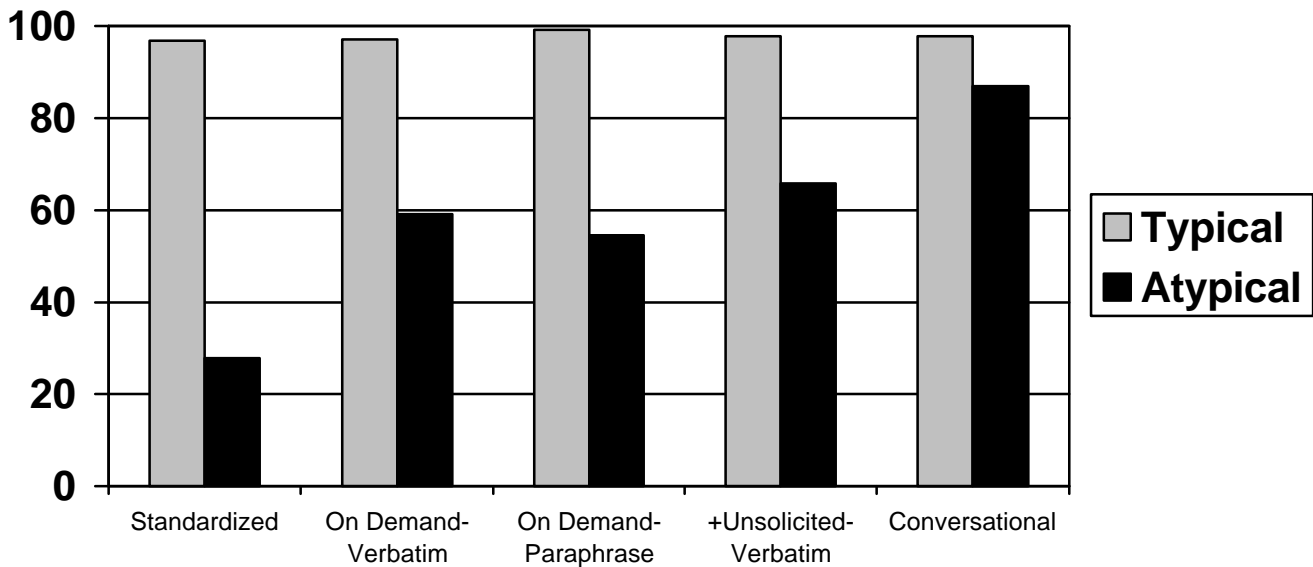
³ We argue elsewhere (Schober and Conrad, in press) that these techniques are not *neutral*. By repeating the question or explicitly leaving the interpretation of the question to the respondent the interviewer suggests that the respondent’s idiosyncratic interpretation is the intended one, which is usually not the case.

In another laboratory study (Schober, Conrad and Fricker, 1999) we tested three additional techniques for providing clarification. These gave the interviewers levels of flexibility somewhere between the level in the two techniques tested in the first study. In two of these additional techniques, interviewers could give help only when it was explicitly requested. In one of these, they had broad discretion in how to present definitions (paraphrased or verbatim, all or part). In the second, they had to provide verbatim definitions (at least one sentence). Finally, in the third technique, interviewers could help whenever it seemed appropriate but they too had to provide verbatim definitions (at least one sentence).

Major Findings: Data Quality and Survey Cost

Respondents answer accurately about typical situations, whether or not interviewers are licensed to clarify the concepts in the questions. However, when respondents answer about atypical situations, their accuracy depends on whether or not they are able to get clarification from the interviewers. When they cannot get this help, they are very inaccurate; when they can obtain it their accuracy improves. The more flexibly the interviewers can provide help, the more it improves response accuracy. Figure 1 presents response accuracy for the two interviewing techniques tested in Schober and Conrad (1997) and the three techniques evaluated in Schober, Conrad and Fricker (1999).

Figure 1. Response accuracy for five interviewing techniques



The first thing to notice is that response accuracy for typical situations (such as the receipt for an end table) is nearly perfect, regardless of the interviewing technique. It does not matter how or when clarification is provided because it is generally not needed. However, when the situations in the scenarios are atypical (as with the floor lamp receipt)

the way help is provided has a large impact on response accuracy. For the Standardized technique, in which there is no way for interviewers to directly provide help, respondents answered correctly only 28% of the time. In contrast, for the Conversational technique, in which interviewers had maximum flexibility in providing help (whenever they deemed it would help, verbatim or paraphrased), respondents provided the right answer 87% of the time. When interviewers had intermediate levels of flexibility (the three middle techniques in Figure 1) respondents' level of accuracy is in between the levels for strictly standardized and conversational interviewing.

Looking more closely at the interaction between interviewers and respondents enriches the picture that emerges from accuracy data alone (see Schober and Conrad, 1997 and Schober, Conrad and Fricker, 1999 for a more detailed discussion). First, interviewers used all five techniques largely as instructed. Second, the two groups of interviewers who were required to present verbatim definitions (On Demand-Verbatim and +Unsolicited-Verbatim) did so almost all of the time that they provided definitions and more than the two groups who were permitted to paraphrase (On Demand- Paraphrase and Conversational); these latter groups almost always provided accurate information even though they sometimes provided it in their own words. Finally, the two groups of interviewers who were permitted to give help when they deemed it appropriate (+Unsolicited-Verbatim and Conversational) frequently did so when it was necessary (in atypical situations) but they sometimes gave help when it was probably not necessary (in typical situations). This last point begins to hint at a potential inefficiency in conversational approaches to clarifying concepts.

Enabling interviewers and respondents to make sure they understand each other clearly helps response accuracy but it also takes time. The duration of the conversational interviews was more than three times that of strictly standardized interviews and the duration of the interviews conducted with intermediate levels of flexibility was more than twice that of the standardized interviews. In fact, there is a strong linear relationship ($r=.98$) between interview duration and response accuracy. For each additional minute that interviewers and respondents spent collaborating, there was a 7% gain in accuracy.

Response accuracy for atypical situations can, essentially, be purchased with the additional time required to converse about question meaning. Such an investment may be worthwhile if the need for accuracy is high. Of course, if it were possible to determine that respondents' situations were almost always typical, then one could be confident of accurate responses without the extra time required for conversational techniques. But one rarely knows this in advance. In these laboratory experiments, half of the scenarios described typical situations and half described atypical situations. In our field experiment (Conrad and Schober, 1999), in which respondents answered about their own lives, we found that their circumstances were atypical frequently enough so that conversational interviewing made a substantial difference in accuracy.

In this study, we used two indirect measures of response accuracy: response change between two interviews and respondents' explanations for their answers. The first interview was standardized for all respondents. The second interview was standardized

for half of the respondents and conversational for the other half. (The techniques were equivalent to the corresponding techniques in the first laboratory study.) If respondents' situations were atypical but they were not able to work this out with the interviewers (as would be the case in a strictly standardized interview) we hypothesized that they would be more likely to misinterpret the question than if they *could* get help (as would be likely in a conversational interview). If respondents participated in two standardized interviews, then their initial misconceptions would likely persist through the second interview and their responses would be consistent between interviews. But in a second interview that is conversational, respondents would be more likely to correctly interpret the question which could lead them to respond differently than they did in the first interview, when they were more likely to have misunderstood.

Responses changed 11% more between a standardized and conversational interview (22% response change) than between two standardized interviews (11% response change). We attribute this difference, at least in part, to respondents' atypical circumstances. If respondents answered "yes" to a purchase question, they were asked to list the purchases on which they based their affirmative response. In the first interview, 57% of respondents' purchases were consistent with the definitions. The same proportion of reported purchases were consistent (57%) when the second interview was standardized. This means that without the opportunity to clarify whether the concepts included their particular purchases, respondents were frequently (43%) incorrect in including their purchases. Surely this would not have happened if their circumstances were typical. When the second interview was conversational and they were able to obtain this clarification, a far greater proportion (95%) of their explanations were legal.

Consider this pair of excerpts from a single respondent in two interviews. In the first (standardized) interview, the interviewer leaves the interpretation of the question up to the respondent⁴:

- (1) I: And how many people live in your home?
R: Five.
I: Five, okay.
R: Well.
I: Uh-huh.
R: We have a, I have a sister that's also in college.
I: Okay, so how so what would you consider.
R: She lives here during the summer and then, she she lives here.
I: Okay so how many would you *say live*.
R: *Five*.
I: Five. Okay.

⁴ In the transcribed excerpts, overlapping speech is enclosed in asterisks. Question marks indicate rising intonation, and utterance-final periods indicate falling or flat intonation, regardless of whether the utterance is a question or an assertion.

The respondent's circumstances are atypical in that it is unclear whether a family member living at school some of the year should be counted as living at home. In the second (conversational) interview, a different interviewer helps this same respondent apply the concept of "live in your home" to her atypical situation and, as a result, the respondent answers differently than in the first interview:

(2) I: Okay, now, how many people live in your home?

R: Five.

I: Five people, okay. And uh uh we uh when I say how many people live in your home what I mean by this is, I don't want you to include uh people that are overnight lodgers, guests, visitors or people who are living away on business, in the armed forces or attending school.

R: Okay then four. We have one that's in school.

I: That's in school okay. So that's four people living in your home. Okay.

Once again, the increase in response accuracy required more interviewing time. The conversational interviews were 80% longer than the standardized interviews (either first or second). While this is shorter than the 300% increase that we observed in the first laboratory study, it is also the case that frequency of atypical situations was lower here than in the laboratory study. Again, it seems that clarification takes time.

Do respondents ask for help when they need it? We have illustrated that respondents can more accurately answer survey questions when they collaborate with interviewers to determine how the survey concepts correspond to their circumstances. The opportunity to collaborate, however, does not guarantee that interviewers and respondents will, in fact, collaborate. On Demand techniques require that respondents first recognize they might need clarification and then explicitly request it. Techniques that also allow interviewers to volunteer clarification require them to first recognize respondents' uncertainty or misinterpretation and to then intervene. There are obstacles to both.

In the first laboratory experiment (Schober and Conrad, 1997), conversational respondents were instructed about the technique. They were told that they might be asked about concepts with special technical definitions and that in order to understand the questions – and to respond accurately – they should ask for clarification. They followed these instructions. When any kind of clarification was provided for atypical situations, it came at the respondents' request 43% of the time; it was volunteered by the interviewers (without an explicit request) 57% of the time. Respondents in the field study (Conrad and Schober, 1999), were given similar instructions but asked for help far less often:⁵ these respondents asked for help on 4% of the occasions that it was provided; the interviewers volunteered the help 96% of the time that they provided it.

This might have happened because, in the field study, respondents' circumstances were typical more than atypical and so they were rarely confused, or it may have occurred because interviewers often (44% of the time) provided definitions before the respondent had a chance to say anything, preempting potential help requests. But there may be

⁵ This analysis was carried out on a subsample of interviews.

another factor involved. Respondents may have been unaware of their need for clarification or unwilling to obtain it.

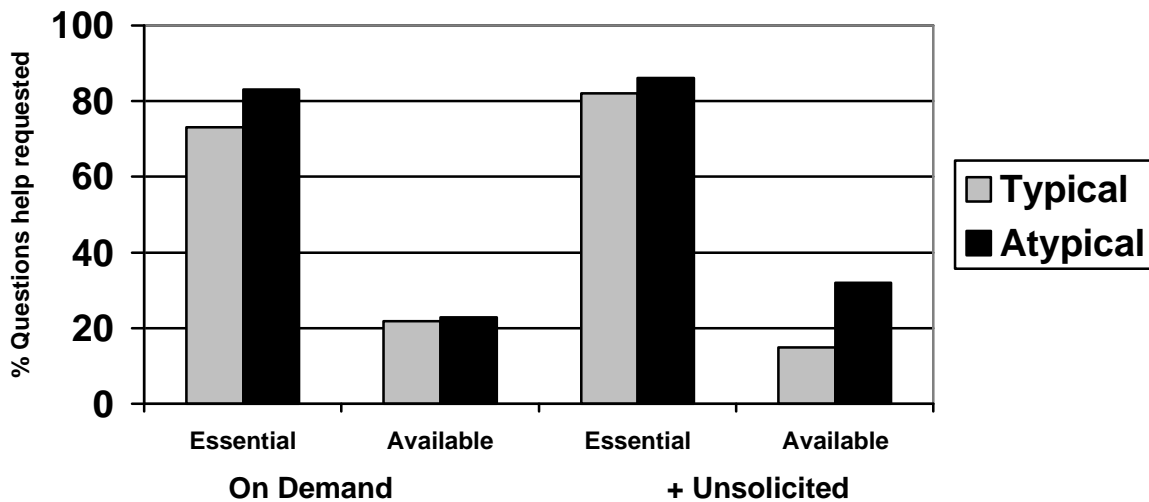
It is possible to confidently misunderstand a question and, thus, be unaware of one's misconception. It is probably hard for interviewers to detect this and so these situations seem destined to contribute to measurement error. There may be other situations in which the respondent is aware of his uncertainty but finds it too costly to ask for help. This cost may be social in origin – it is embarrassing to acknowledge some confusion about an everyday term – or the cost may be cognitive – it requires mental effort to formulate a question that is likely to elicit the clarifying information. The respondents in the field study might have been less willing than their laboratory counterparts to bear these costs. Is there some way to lower these costs and promote more clarification?

In another laboratory study (Conrad and Schober, 1998), we attempted to lower the costs of obtaining help by implementing three interviewing techniques as Computer Assisted Self Interview (CASI) instruments. In our Standardized implementation, the question appeared on the computer screen and the respondent typed in the answer; there was no way to clarify meaning. In an On-Demand instrument, respondents could click their mouse on highlighted text to display a textual definition. In the +Unsolicited implementation, respondents could also click their mouse on the highlighted text or, if they were inactive for more than a fixed amount of time, the computer would offer them a definition. In all cases respondents answered on the basis of the same scenarios used in the other laboratory studies.

Two groups of respondents interacted with the On-Demand and +Unsolicited instruments. One group was told that they could obtain definitions if they chose to by clicking the mouse (Definitions Available); the other was also told that, in some cases, they would have to obtain definitions in order to respond accurately (Definitions Essential).

We hypothesized that the CASI implementations would lower the social costs of asking for help because they made it possible to do so privately (Tourangeau and Smith, 1998). We reasoned that CASI would also lower the cognitive costs of obtaining clarification because, instead of formulating a question that articulates her uncertainty, the CASI respondent only needs to click a mouse. While we cannot tease these costs apart in the current study, it is clear that some respondents were very willing to ask the computer for definitions (see Figure 3). Respondents who understood the importance of asking for help (Definitions Essential) did so frequently (81% of the time); those for whom this was not emphasized (Definitions Available) were relatively unlikely to ask for help (23% of the time). This pattern was unrelated to the typicality of the situations. We interpret this as evidence that obtaining help was sufficiently low cost that, once respondents appreciated its importance, they clicked for a definition for almost all questions, even if they had no sense of uncertainty.

Figure 2. Percent questions on which help requested



Conclusions

Conversational interviewing is an option available to survey researchers. It is not a panacea. While the technique can improve response accuracy in some circumstances, there are costs associated with its use. The clearest cost is interview duration. It takes time to clarify concepts. It may be that clarification becomes streamlined as conversational interviewers become experienced, providing just the information that is relevant to a respondent's misconception or uncertainty. But even this takes more time than no clarification. And there are costs associated with training interviewers on the concepts and defining the concepts in the first place.

Yet these costs may be worth it. Conversational interviewing leads to more consistently correct question interpretations than does standardized interviewing, particularly when respondents' situations are atypical. And atypical situations are frequent enough, that response accuracy suffers substantially when respondents cannot obtain help in standardized interviews. Conversational interviewing will produce the greatest payoff when respondents ask for the clarification they need. In the end, the choice of method depends on what researchers value more: ease of administration and low cost or confidence that responses are accurate.

References

- Beatty, Paul. 1995. "Understanding the Standardized/Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11:147-160.
- Cicourel, Aaron V. 1973. *Cognitive Sociology: Language and Meaning in Social Interaction*. New York: The Free Press.
- Clark, Herbert H. 1992. *Arenas of Language Use*. Chicago: University of Chicago Press.

- Conrad, Frederick G., and Michael F. Schober. 1998. "A conversational approach to computer-administered questionnaires." In *Proceedings of the American Statistical Association, Section on Survey Methods Research*. Alexandria, VA: American Statistical Association.
- Conrad, Frederick G., and Michael F. Schober. 1999. "Conversational interviewing and measurement error in a household survey." Manuscript under review.
- Fowler, Floyd J. and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: SAGE Publications, Inc.
- Kovar, Mary G. and Patricia Royston. 1990. Comment on Suchman and Jordan." *Journal of the American Statistical Association* 85 (409): 246-247
- 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, UK: Cambridge University Press.
- Schober, Michael F., and Frederick G. Conrad. 1997. "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61:576-602.
- Schober, Michael F., and Frederick G. Conrad. 1998. "Response Accuracy When Interviewers Stray from Standardization." In *Proceedings of the American Statistical Association, Section on Survey Methods Research*. Alexandria, VA: American Statistical Association.
- Schober, Michael F., Frederick G. Conrad and Scott S. Fricker. (1999, May). "Further explorations of conversational interviewing. How gradations of flexibility affect cost." Paper presented at the 54th Annual Conference of the American Association for Public Opinion Research, St. Petersburg, FL.
- Schober, Michael F., and Frederick G. Conrad. In press. "A Collaborative View of Standardized Survey Interviews." In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, ed. Douglas Maynard, Hanneke Houtkoop, Nora Cate Schaeffer, and Johannes van der Zouwen. New York: Wiley.
- Suchman, Lucy, and Brigitte Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85 (409):232-253.
- Tourangeau, Roger. 1990. "Comment on Suchman and Jordan." *Journal of the American Statistical Association* 85 (409):250-251.
- Tourangeau, Roger and Tom W. Smith. 1998. "Collecting sensitive information with different modes of data collection." In Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nicholls, W. L., and O'Reilly, J. (Eds.) *Computer Assisted Survey Information Collection*. New York: New York: John Wiley & Sons, 431-453.