RESPONSE ACCURACY WHEN INTERVIEWERS STRAY FROM STANDARDIZATION[1]

Michael F. Schober, New School for Social Research
Frederick G. Conrad, Bureau of Labor Statistics
Michael Schober, Dept. of Psychology AL-340, New School for Social Research,
65 Fifth Ave., New York, NY 10003

Key Words: Standardized interviewing,
conversational interviewing, data quality

INTRODUCTION

Survey researchers have long debated what exactly interviewers should do when respondents aren't sure how to interpret survey questions (see Beatty, 1995). The predominant philosophy has been that the best way to maintain objectivity is for interviewers to read questions exactly as worded and to leave the interpretation of questions entirely up to respondents. Under this sort of strictly standardized interviewing (see, e.g., Fowler & Mangione, 1990), if respondents show that they need help interpreting the question or formulating their answers, interviewers must avoid doing anything that biases the response. Consider, for example, the following exchange:

> Interviewer (I): In the past year, that is, since May of 1997, have you purchased or had expenses for household furniture?
> Respondent (R): Um...would you count a lamp as furniture?

At this point, a strictly standardized interviewer could offer to repeat the question, or she might explicitly state the response alternatives ("Would that be a yes or a no?"). But she should never interpret the question for the respondent, nor should she explain what the survey designers mean by particular expressions in the questions like "household furniture"; explanations by interviewers are likely to be given unsystematically (not all respondents will get them), and they are likely to mislead respondents at least as often as they help them out (see Fowler & Magnione, 1990, p. 21). Of course, this approach can work only if questions have been well-pretested, so that questions are easily interpretable.

Under an alternative approach, which we call *conversational interviewing*, interviewers and respondents work together to assure that respondents interpret questions as the survey designers intended. Interviewers, rather than being restricted to repeating the question and probing neutrally, may say whatever it takes to make sure that respondents have interpreted questions appropriately. In our example, the interviewers might explain that, for the purposes of this survey, the sponsoring organization doesn't count lamps and lighting fixtures as pieces of household furniture.

One might think that the simplest solution to dealing with potential misunderstanding is to modify the survey question to include definitional information, and then to administer the survey using strictly standardized interviewing techniques. But this simplest solution turns out to be impractical in many cases. In our example, although the question could easily be modified to deal with lamps and lighting fixtures—"Do not include lamps and lighting fixtures"—the sponsoring organization may have far too many other caveats in their official definition. For example, one government organization defines household furniture 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 applicances. (U.S. Department of Commerce, 1993).

A question that included all this information would probably be very unpleasant, if not uninterpretable, for respondents.

In a recent laboratory study (Schober & Conrad, 1997) we examined response accuracy—the extent to which responses matched the definitions of the sponsoring organizations—in strictly standardized and conversational interviews. In the study, we trained professional telephone interviewers to conduct extreme versions of the two interviewing techniques.

Respondents in our laboratory answered questions from ongoing government surveys from fictional scenarios that we had devised: floor plans, work descriptions, and purchase receipts. Thus we could measure response accuracy directly, since we knew what the "correct" (with respect to the official definitions) answers were.

For each question, we devised two alternate scenarios (see, for example, Figure 1). With one scenario, the survey question should be easy to interpret for all respondents; this scenario led to a *straightforward mapping* between the question and the respondent's (fictional) circumstances. With the alternate scenario, it was less clear how the survey question should be answered; the scenario led to a *complicated mapping* between the question and respondent's circumstances. For any one question, a respondent would see either the straightforward-mapping scenario or the complicated-mapping scenario; in each interview, half the scenarios would lead to straightforward mappings and half to complicated mappings. As Figure 1 shows, for example, for the household furniture question, some respondents saw a receipt for an end table (straightforward mapping), and others saw a receipt for a floor lamp (complicated mapping). The official definitions always clarified what the correct answers should be.

The results showed that interviewing techniques made no difference for straightforward mappings; performance was nearly perfect. For complicated mappings, strictly standardized interviewing led to poor accuracy, while conversational interviewing substantially improved accuracy. But the improvement in accuracy came at a substantial cost—conversational interviews took much longer.

The interviewing techniques implemented in Schober and Conrad (1997) probably differ from actual practice. Even in organizations that train interviewers to be strictly standardized, interviewers deviate from strict standardization some percentage of the time (see, e.g., Bradburn et al., 1979; Brenner, 1982; Mangione et al., 1992, Morton-Williams, 1979). And some organizations encourage their interviewers to provide definitions for respondents whenever respondents request them. The question we address here is how the interviewing techniques actually practiced in a mainstream survey organization affect response accuracy, and whether this differs for straightforward and complicated mappings.

EXPERIMENT

In this experiment, we use exactly the same procedure as in the Schober and Conrad (1997) study to directly examine response accuracy when interviewers do what they ordinarily do. Respondents in the laboratory answered questions from ongoing surveys on the basis of fictional scenarios, so that we could measure the extent to which their answers matched the official definitions of survey concepts.

Questions. Respondents were asked the same questions used in the Schober and Conrad (1997) study. Twelve questions from three surveys were used, four questions about employment from the Current Population Survey (e.g., "Does anyone in this household have a business or a farm?"), four questions about housing from the Consumer Price Index-Housing survey (e.g., "How many people live in this house?"), and four questions about purchases adapted from the
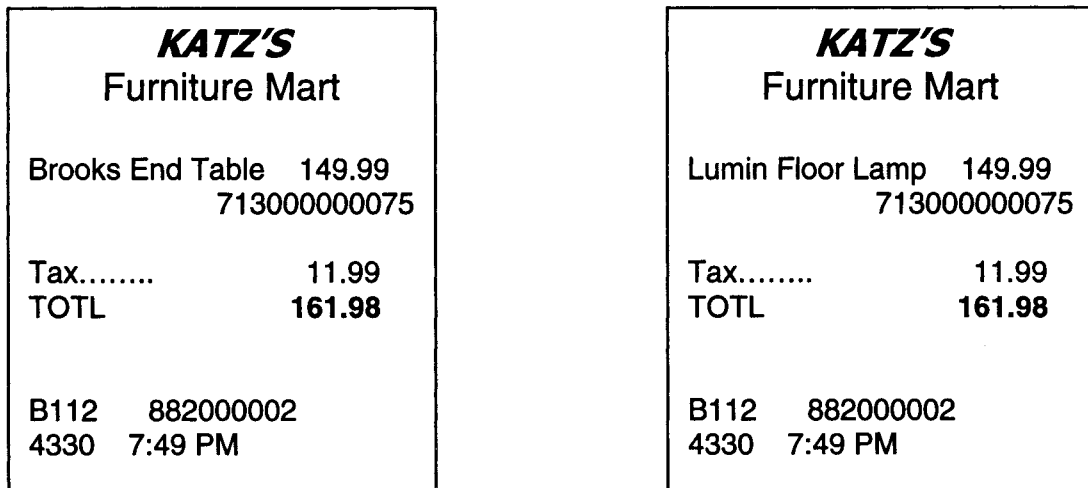
| ***KATZ'S*** |
| Furniture Mart |
| |
| Brooks End Table   149.99 |
| 713000000075 |
| |
| Tax........   11.99 |
| TOTL   **161.98** |
| |
| B112   882000002 |
| 4330   7:49 PM |

| ***KATZ'S*** |
| Furniture Mart |
| |
| Lumin Floor Lamp   149.99 |
| 713000000075 |
| |
| Tax........   11.99 |
| TOTL   **161.98** |
| |
| B112   882000002 |
| 4330   7:49 PM |

Figure 1: Example scenarios, straightforward and complicated mappings

Current Point of Purchases Survey (e.g., "Last week, did Carla purchase or have expenses for car tires?"). The questions had all been pretested for intelligibility. Official definitions existed for key concepts in all the questions.

Stimuli. Respondents answered the questions on the basis of the fictional scenarios—floor plans, work descriptions, and purchase receipts—from the Schober and Conrad (1997) study. Scenarios weren't available to the interviewers, so interviewers did not know what the correct answers to the questions were, and they could not predict the correct answers from previous interviews. As in the earlier study, for each respondent 50% of the scenarios had complicated mappings and 50% had straightforward mappings

Participants. The 11 interviewers were professional Census Bureau interviewers (10 female, 1 male) at the Hagerstown, MD telephone facility. They averaged 20.6 months interviewing experience, ranging from 5 to 45 months. Each interviewer telephoned two paid respondents in the Bureau of Labor Statistics laboratory in Washington, DC (except for one interviewer who only telephoned one respondent).

The 21 respondents were recruited from ads in the Washington Post. They came from a range of demographic backgrounds; 10 were female and 11 were male; 7 were Black, 12 were White, and 2 were Asian; and they ranged in education from high school diplomas to graduate degrees. The demographics of the group were much the same as for the respondents in the Schober and Conrad (1997) study, and so it is reasonable to make comparisons across the groups.

Interviewer training. All interviewers were trained on the key survey concepts, using group discussion and a quiz; the same training was used in the Schober & Conrad (1997) study. Interviewers were then asked to conduct the interviews exactly as they ordinarily do. Note that official practice at the Hagerstown facility is not strictly standardized. Although interviewers are instructed to read question exactly as worded and to use only nondirective probes, they are also allowed to clarify questions at the respondents' request.

RESULTS

Deviation from strict standardization. How strictly standardized were interviewers? Interviewers conformed to strict practice for 80.2% (204) of the 254 question-answer sequences. In the 19.8% of cases where they deviated from strict practice, they deviated more often when mappings between questions and circumstances were complicated (27.0% of cases) than

when the mappings were straightforward (12.7% of cases).

The most frequent deviation was asking unscripted clarification questions (34/50 cases). Consider this example, where after the respondent answered the question ambiguously, the interviewer probed in a way that reveals what the official definition counts as a bedroom.

I: How many bedrooms are there in this house.
R: Originally two, but they're using three
I: Okay but there were two rooms designed specifically for bedrooms?
R: Right.
I: Okay.

In the following example, the interviewer uses follow-up questions to hone in on the various features of the official definition, and essentially tells the respondent what the answer ought to be:

I: Does anyone in this household have a business or a farm.
R: Um, Harry does, but he ju- I mean- just works as a gardener though.
I: Okay so is it it's um like a business of his?
R: Uh he just does it uh on the side ever since he retired?
I: Okay, so um like so um do you classify is a- as a business I mean does he have like a listing in a- in a classified section in the telephone or *is it on the side*
R: *No, no*, it's just something that he does on the side.
I: So it's- I mean so it's technically it's not really a business then is that what you're saying, it's just something he *does-*
R: *It's* not a business, but he did- did
I: Is *this-*
R: *consider it* work, I'd consider it work though
I: Okay, but not a business?
R: Yes.
I: Okay.

In addition to using unscripted follow-up questions, interviewers deviated from strict standardization by providing the official definitions (either reading them verbatim or paraphrasing), rewording questions in order to probe further, and explicitly telling respondents what the answer should be, as in "That would be two bedrooms, then".
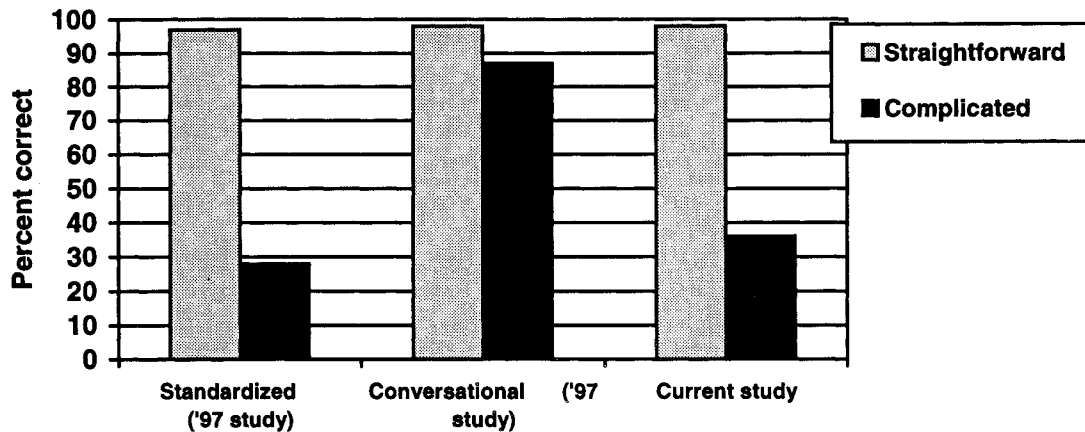
**Figure 2: Response accuracy in current study compared with Schober and Conrad (1997)**

Most of these deviations from strictest practice conformed with the spirit (if not the letter) of what interviewers had been trained to do; recall that they are encouraged to provide definitions when respondents request them. And the information they provided was highly accurate; 45 of the 50 deviations from strict standardization were entirely accurate and could not be seen as misleading the respondents.

Response accuracy. Respondents' answers matched the official definitions virtually perfectly for straightforward mappings (98% accuracy). But their response accuracy for complicated mappings was quite poor, 36%. If we compare these accuracy rates to those in the Schober and Conrad (1997) study (see Figure 2), we see that the pattern is not reliably different than for the strictly standardized interviews in that

study, and that accuracy is far poorer than in conversational interviews. This demonstrates, on the one hand, that interviewers at this organization are performing comparably to strictly standardized interviewers; unfortunately, their performance does not lead to high levels of response accuracy for complicated mappings.

A more detailed examination of response accuracy shows that interviewer deviations from strict standardization led to substantial improvement in accuracy (see Figure 3). In fact, the 79% accuracy rate for complicated mappings when interviewers deviated is comparable to the 87% rate in our earlier conversational interviews.

This finding is supported from several other views of the results. If we examine average response accuracy
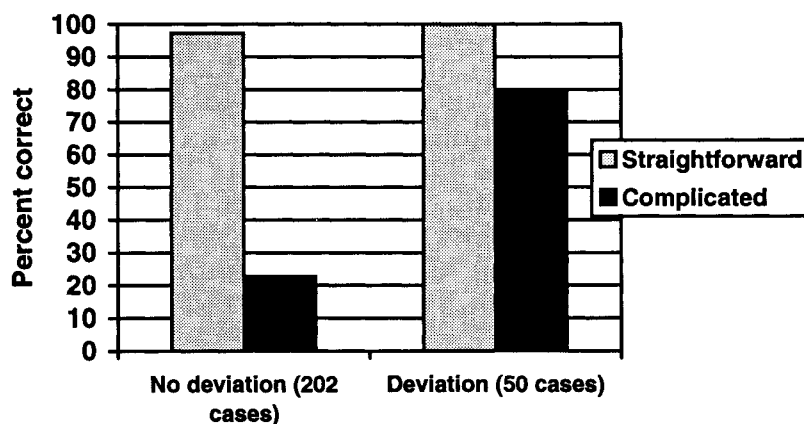
Figure 3: Response accuracy when interviewers deviated from strict standardization

for each interviewer, we see that interviewers who deviated from strict standardization the most drew the most accurate responses for complicated mappings, $r(11) = .76$, $p = .007$. If we examine how often interviewers deviated from strict standardization on a question-by-question basis, we see that response accuracy for complicated mappings was greatest for those questions on which interviewers deviated the most, $r(12) = .59$, $p < .05$. Of course, the sample is quite small and we should be careful about generalizing, the but the overall picture is consistent: greater deviation from strict standardization can lead to improved response accuracy for complicated mappings.

Interview duration. Interviews took a median of 4.26 minutes, ranging from 2.95 to 8.18 minutes. They did not take reliably longer than the standardized interviews in the Schober and Conrad (1997) study (3.41 minutes, ranging from 2.48 to 5.99); both standardized interviews and the interviews in the current study were much quicker than the conversational interviews (11.47 minutes, ranging from 6.10 to 35.44 minutes). Examination of the total number of words used by interviewers and respondents when mappings were straightforward and complicated shows the same pattern (see Figure 4): interviews in the current study did not use reliably more words than the standardized interviews in the earlier study, but they used far fewer words than conversational interviews.

If we break down these word counts further, we see that what really took time was interviewers' deviations from strict standardization (see Figure 5). For those
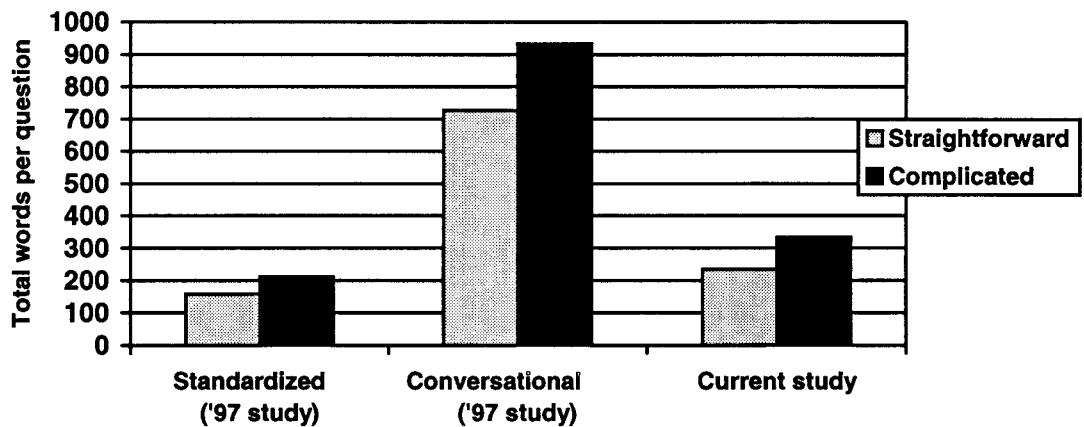

Figure 4: Interview length in current study, compared with Schober and Conrad (1997)
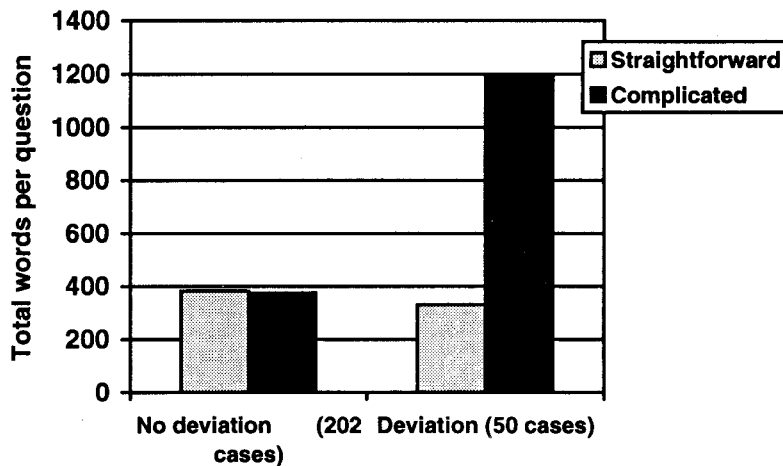

Figure 5: Interview length when interviewers deviated from strict standardization

complicated mapping cases where interviewers didn't deviate from strict standardization, the interviewers took no longer than for straightforward cases.

So the general picture is that response accuracy can improve substantially when interviewers deviate from strict standardization, but this deviation takes time. This provides more naturalistic evidence for the tradeoff between accuracy and efficiency noted in Schober and Conrad (1997) and Conrad and Schober (under review).

Interestingly enough, interviewers with the most experience (measured in months) produced the least accurate responses for complicated mappings, $r(11) = -.63$, $p = .038$. This may be related to the (nonsignificant) trend for more experienced interviewers to deviate from strict standardization less often, $r(11) = -.51, p = .112$.

SUMMARY AND CONCLUSIONS

These results demonstrate, first, that interviewers at this organization conducted interviews in just the way they were supposed to. Unfortunately, overall this led to poor response accuracy for complicated mappings. Response accuracy improved substantially in those cases where interviewers strayed from strict standardization in order to make sure that respondents were answering according to official definitions.

These data provide more naturalistic support for the findings in Schober and Conard (1997), where interviewers were trained to conduct interviews in ways they ordinarily don't, while still allowing us to directly measure the fit between respondents' answers and the official definitions. As in that study, here there were clear benefits for deviation from standardization: improved response accuracy for complicated mappings. Also as in that study, these benefits came at a real cost: deviation from standardization increased interview length, for those questions where interviewers deviated.

To the extent that the practices in this interviewing facility are typical, our results show that training procedures already empower interviewers to use some conversational interviewing techniques (providing definitions when respondents ask for them). Interviewers also seem to use some unlicensed techniques (e.g., telling respondents what the answer should be when respondents describe their circumstances). Even the unlicensed techniques seem to help respondents answer questions in ways more closely aligned to official definitions, as Suchman and Jordan (1991) have argued. All in all, current practice is far closer to strictly standardized than to conversational interviewing, and as such it leads to poorer response accuracy than it might.

REFERENCES

Beatty, Paul. 1995. "Understanding the Standardized/ Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11:147-160.

Bradburn, Norman M., et al. 1979. *Improving Interview Method and Questionnaire Design.* San Francisco: Jossey-Bass.

Brenner, Michael. 1982. "Response Effects of Role-Restricted Characterstics of the Interviewer," in W. Dijsktra, and J. van der Zouwen (eds.), *Response Behavior in the Survey Interview*, pp. 131-165, London: Academic Press.

Conrad, Frederick G., and Michael F. Schober. "Conversational Interviewing Can Reduce Survey Measurement Error: Evidence from a Household Telephone 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.

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

Morton-Williams, J. 1979. "The Use of 'Verbal Interaction Coding' for Evaluating a Questionnaire," *Quality and Quantity* 13:59-75.

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

Suchman, Lucy, and Brigitte Jordan. 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. 241-267. New York: Russell Sage Foundation.

U.S. Department of Commerce. 1993. *Point of Purchase Survey 1993, Checklist A: Five-Year to One-Week Recall* (CPP-2A). Washington, DC: Bureau of the Census.