

HOW INTERVIEWERS' CONVERSATIONAL FLEXIBILITY AFFECTS THE ACCURACY OF SURVEY DATA¹

Frederick G. Conrad, Bureau of Labor Statistics
Michael F. Schober, New School for Social Research
Frederick Conrad, BLS, 2 Massachusetts Ave. NE, Room 4915, Washington, DC 20002

Key Words: Standardized interviewing,
conversational interviewing, response accuracy

INTRODUCTION

The prevailing philosophy of survey data collection—standardization—requires interviewers to use identical words and to probe nondirectively. Standardization, in some form, has been widely practiced for some time; recently Fowler and Mangione (1990) have articulated an explicit rationale and procedures for its use. An alternative perspective, which we call *conversational flexibility*, gives interviewers license to deviate from standard scripts in order to help respondents understand questions in the way survey designers intended them. Conversational flexibility has been championed in the survey world by Suchman and Jordan (1990, 1991), and by researchers in other disciplines (see Holstein & Gubrium, 1995; Kvale, 1994; Mishler, 1986; and others). Advocates of both standardization and conversational flexibility claim greater accuracy for their method, but as Schaeffer (1991, p. 371) puts it, "...it is an open question whether systematically giving participants more access to 'normal' conversational resources will improve the quality of the interaction or the resulting data" (see also Beatty, 1995). Which method actually leads to superior response accuracy?

STANDARDIZATION VS. CONVERSATIONAL FLEXIBILITY

In a standardized interview, the interviewer presents exactly the same stimulus to all respondents: The interviewer always reads exactly the same question and never interprets the question for the respondent. If the respondent solicits help (or the interviewer otherwise believes the respondent needs help), the interviewer may only probe "neutrally": repeat the question, repeat the response alternatives, ask for the respondent's interpretation, etc. The idea is that the stimulus—the words uttered by the interviewer—should be consistent from one interview to the next.

Proponents argue that standardization will lead to superior accuracy primarily because the interviewer doesn't have the opportunity to mislead the respondent. As long as questions have been well pretested, question comprehension should be adequate when the questions are presented uniformly.

In a conversationally flexible interview, the interviewer and respondent work together to assure that the respondent interprets the question as the survey designer intended. After asking the question, the interviewer may clarify it as needed, and say whatever it takes to help the respondent interpret the question as intended.

Proponents believe flexible interviews will lead to superior response accuracy because respondents do not always understand questions as the designers intend them to be understood, even if the questions have been pretested extensively. Different respondents may reach different interpretations, and at least some of those may be wrong. The philosophy is that in ordinary communication, the desired understanding is only guaranteed through conversational collaboration (Clark & Schober, 1991; Clark & Wilkes-Gibbs, 1986). Survey interviews are specialized conversations, and participants should be able to collaborate to understand each other just as they do in ordinary conversation. As a result, what should be standardized is not *wording*, but rather *meaning* (Suchman & Jordan, 1991).

One possibility is that each approach is effective in different situations. When the concepts in the question clearly correspond to the respondent's life circumstances, standardized interviewing may be effective. In contrast, when this correspondence is ambiguous, more respondents may be able to answer the question as intended if interviewers can clarify questions. This may improve overall response accuracy.

OVERVIEW OF STUDY

We carried out a laboratory experiment to compare the accuracy produced by flexible and standardized

¹We thank Clyde Tucker and Sarah Jerstad for help in various aspects of the project, and Susan Brennan and Brian Kojetin for comments on an earlier draft. We also thank ASA/NSF/ BLS for fellowship support of Michael Schober. The views expressed here reflect the opinions of the authors and not those of the Bureau of Labor Statistics.

interviewing techniques. This required that we set up a situation in which we knew the true values (the correct responses) for all the questions asked. Rather than asking respondents about their own lives we asked them about fictional scenarios that we created. This allowed us to establish the correct answer according to published definitions.

We trained telephone interviewers to implement either technique in a “pure” form. (The philosophy of standardized interviewing is widespread, but in practice many survey interviewers use some combination of standardized and flexible interviewing. We used pure versions of both techniques so that we could see the effects most clearly). Our interviewers asked questions from ongoing government surveys.

METHOD

Design. The experiment involved administering either standardized or flexible interviews to different groups of respondents. In each interview, the interviewer asked questions about fictional scenarios. Half of the questions corresponded to the scenarios in a straightforward way and half corresponded in a complicated way. All respondents were asked the same questions; what differed was the type of interaction in the interview.

Participants. The “respondents” were experimental subjects recruited from the subject pool maintained by the Bureau of Labor Statistics. Forty-one respondents were paid \$25.00 each for one experimental session. They varied in their demographic backgrounds (22 women, 21 men; 14 Black, 25 White, 2 Asian; educational backgrounds ranging from high school to professional degrees). We tried to balance the demographic characteristics of the two interviewing groups, so that roughly equal numbers of high-school-educated participants were interviewed with standardized and flexible techniques, etc. Beside this balancing, respondents were assigned to interviewing conditions at random.

Twenty-two Census Bureau interviewers (17 women, 5 men) at the Hagerstown, Maryland telephone facility each called two respondents in the BLS laboratory (except for one interviewer who only called one respondent) for a total of 43 interviews.

Questions. There were 12 questions, each from a current, major government survey. Four were drawn from the Current Population Survey (CPS), four from the Consumer-Price Index Housing survey (CPI-Housing), and four from the Current Point of Purchases Survey (CPOPS), which is part of the Consumer Price Index program. All these questions had been pretested; those from CPS had been extensively pretested. The questions dealt with the

domains of employment, housing, and purchases respectively. Official definitions existed for key concepts in all questions. These had been developed and published by each survey program

Scenarios. As we indicated before, respondents answered on the basis of fictional scenarios for which we knew the correct answers according to official definitions. The scenarios included floor plans, purchase receipts, and textual vignettes. These 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 scenarios. The way knowledge was allocated to the participants was therefore analogous to its division in an actual survey: The respondents knew the “facts” and the interviewers knew the questions and concepts.

The way the questions read by interviewers and the respondents’ scenarios corresponded with each other was pivotal. For each interview, half these “mappings” were *straightforward* and half were *complicated*. The official survey definitions on which the interviewers had been drilled always clarified the mappings. This was particularly important in the complicated cases.

Consider this question taken from the CPI-Housing survey: *How many half bathrooms are there in this house?* In our study, a respondent who had a scenario that led to a straightforward mapping would see the floor plan in Figure 1a. The room at the lower left is unambiguously a half bathroom—it has a toilet and sink. So in this sense the mapping to the question is straightforward. In contrast, a respondent whose scenario led to a complicated mapping would see the floor plan in Figure 1b, where the room in the lower left has only one fixture. A respondent faced with this floor plan might wonder whether this room should be considered a half bathroom. In fact the official CPI-Housing definition requires that half bathrooms have two fixtures (either a toilet and a sink, a sink and a shower/tub, or a toilet and a shower/tub), and so this would not be considered a half bathroom. Without knowing the CPI-Housing definition, a respondent might not know how to respond.

Training. All interviewers were trained on the key survey concepts for about one hour. Eleven were selected at random and trained for an additional hour to conduct a pure version of standardized interviews where they read questions exactly as worded and provided only nondirective probes, but never provided definitions for the survey concepts. Probing techniques

included rereading the question, providing the response alternatives, and otherwise probing neutrally.

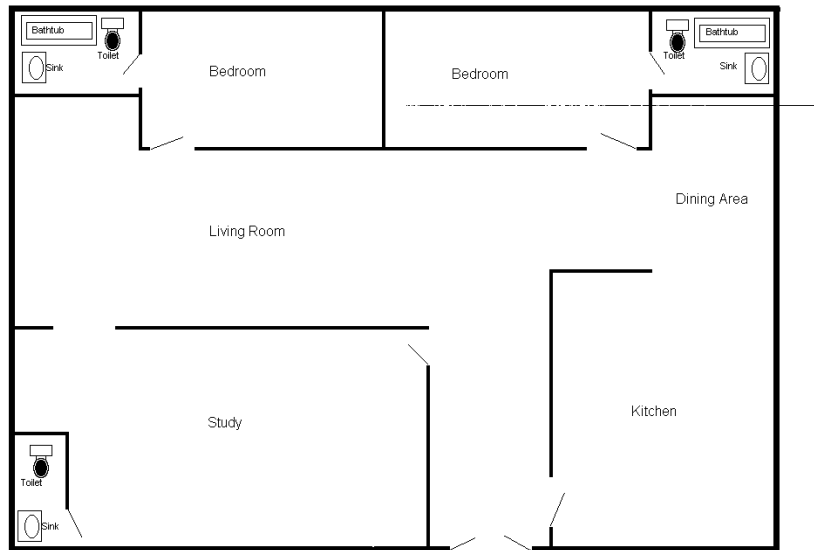


FIGURE 1a: Example scenario designed to create straightforward mapping. Room in lower left corner has two fixtures, and so officially qualifies as a half bathroom.

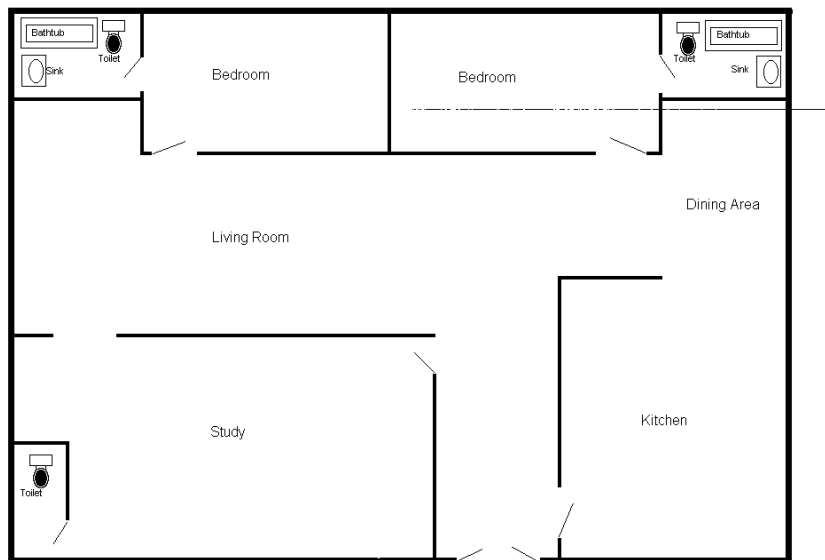


FIGURE 1b: Example scenario designed to create complicated mapping. Room in lower left corner has only one fixture, and so officially does not qualify as a half bathroom.

The other 11 interviewers were trained for an hour to conduct flexible interviews. Like standardized interviewers, these interviewers were instructed to read the questions exactly as worded, 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.

Instructions to Respondents. Respondents in flexible interviews were encouraged to ask for clarification as necessary. For the technique to be effective, participants must work together in conversation to understand each other, so we encouraged respondents to ask questions where they needed help. Respondents receiving standardized interviews received no special instructions as is the case in typical standardized surveys.

RESULTS

Before we turn to the accuracy results, we first present some general characteristics of the interactions. These suggest that the interviewers correctly implemented each interviewing technique. First, standardized interviews contained a high proportion of *simple question-answer sequences*. These are sequences in which the interviewer asked the question exactly as worded and the respondent immediately provided an answer, followed by no other substantive interactions. About 70% of the question-answer sequences were simple (a question followed immediately by the answer) in standardized interviews, compared with only 10% in the flexible interviews. Second, we coded *directive interventions*—all those interventions that would be “illegal” in pure standardized interviewing. Those occurred in 85% of

the questions in flexible interviews, but in only 2% of the questions in standardized interviews. (This is a liberal count; the 2% “illegal” interventions in standardized interviews were all incomplete repetitions of the question, and so by some counts would be legal). So clearly the two types of interviews were implemented in qualitatively different ways, and much as we had intended.

Now turning to the accuracy results: In the case of standardized interviews, when the mapping between the question and the scenarios was straightforward, response accuracy was nearly perfect, about 97% (see Figure 2). But when the mapping was complicated—that is, when a key concept was required to make it clear—accuracy dropped to about 28%. Just as in the standardized interviews, flexible interviews also led to nearly perfect response accuracy (about 98%) when the mapping was straightforward. But unlike in the standardized interviews, accuracy was also very high when the mappings were complicated, 87%. Looking just at complicated mappings, accuracy was nearly 60 percentage points higher with flexible interviewing than with standardized (interaction of scenario complexity x interview type: $F[1, 41] = 130.01, p < .001$, by subjects; $F[1,11] = 100.74, p < .001$, by items). This is not at all what standardization theory would predict.

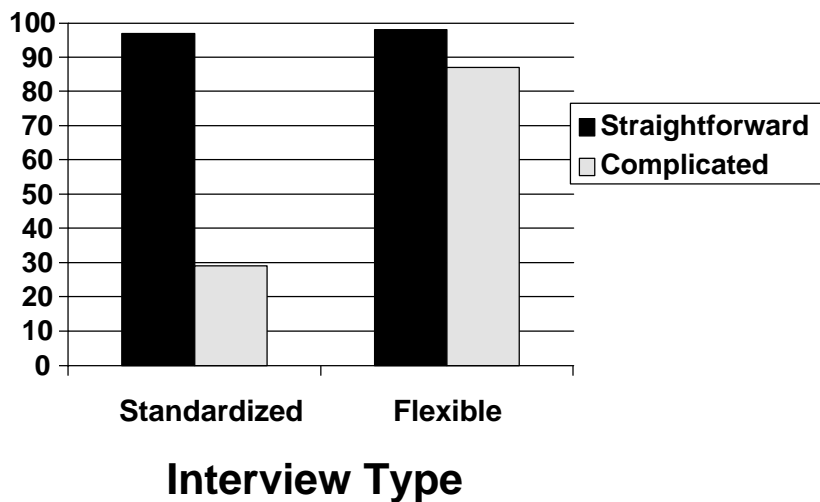


FIGURE 2: Response accuracy (percentage)

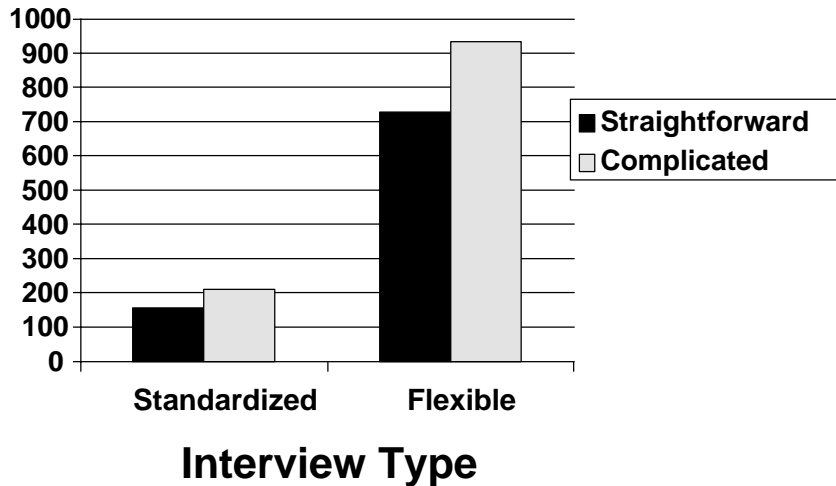


FIGURE 3: Total words per question by both interviewer and respondent

One concern that proponents of standardized interviewing have raised about flexible interviewing is that interviewers can mislead respondents. That is, even if interviewers sometimes provide information which helps respondents to produce accurate answers, interviewers may just as often provide information which can lead respondents astray. For 2/3 of the observations, flexible interviewers provided some sort of explicit, directive information. In 87% of these cases, interviewers provided accurate information and respondents provided accurate answers. In six percent of the cases, interviewers provided accurate information but respondents answered incorrectly. On four percent of occasions interviewers provided some inaccurate information but respondents still produced the correct answer. And three percent of the time interviewers provided some inaccurate information and respondents answered incorrectly. So flexible interviewers generally provided highly accurate information, and when they provided inaccurate information this didn't necessarily lead respondents to produce incorrect answers.

Not only did flexible interviewers provide mostly accurate information, they provided most of this information when it was most useful, that is when mappings were complicated. This occurred even though interviewers couldn't see the respondents'

scenarios, and never knew at the outset whether respondents were faced with complicated or straightforward mappings. With complicated mappings interviewers provided information most of the time (90%). Some of these times respondents had explicitly requested help (39%), and sometimes the help was unsolicited (51%). With straightforward mappings, on the other hand, respondents rarely asked for help (1%) and interviewers rarely volunteered any (2%) (see Schober & Conrad, in preparation, for further details).

Although flexible interviewing led to massive improvements in accuracy, and although flexible interviewers rarely misled the respondents, the technique does have a significant cost: It takes much longer. The median time to complete standardized interviews was 3.41 minutes, compared to 11.47 minutes for flexible 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 Figure 3 shows, flexible interviews took longer than standardized interviews, $F(1,40) = 64.24, p < .001$, and complicated scenarios took slightly longer than straightforward ones, $F(1,40) = 7.31, p = .01$. There was no interaction.

So there is a clear tradeoff between improved accuracy and increased time. But this tradeoff may be

less extreme than it seems. First, our interviewers were new to the definitions. As a result, some flexible interviewers were not adept at focusing on just the relevant parts of definitions, and read entire lengthy definitions verbatim. More experienced flexible interviewers might be better at tailoring their explanations, which would lead to shorter flexible interviews. Second, our experiment implemented pure versions of the interviewing techniques. The common wisdom is that actual practice often involves some combination of both techniques, which undoubtedly leads to longer interviews than pure standardized techniques would produce. Flexible interviewing might not increase interview length threefold (as in our experiment) when compared with interviews as currently practiced.

SUMMARY AND CONCLUSIONS

In this study, there were clear benefits for flexible interviewing: a 60% increase in 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.

But there were real costs in administering flexible interviews: a more than threefold increase in duration (though, as we have noted, our flexible interviews may be longer than they would be under field conditions, and our standardized interviews may be shorter than ordinary "standardized" interviews).

To evaluate whether flexible interviewing is worth the investment, we need to know how often the mappings between questions and respondents' circumstances are complicated within a survey's domain. If the mappings are often complicated, or if it is not known whether the mappings are complicated or straightforward, our data argue that the extra cost of flexible interviewing may be worthwhile. We recommend the use of standardized interviewing only if practitioners can demonstrate that the mapping between their questions and respondents' situations is straightforward. If practitioners cannot demonstrate this, they need to recognize the potential loss of response accuracy resulting from standardized interviewing.

Our data also allow us to speculate about various possible implementations of flexible interviewing. Under the simplest implementation, one that would require the least revision of current practice (and in fact, may conform to some current practice), interviewers could be licensed to provide scripted definitions only when respondents request help. While our data suggest that this would lead to an increase in accuracy, accuracy improved even more when

interviewers also intervened without respondents' requesting help. In our experiment, the more interviewers used their knowledge and discretion, the more accurately respondents answered the questions.

REFERENCES

- Beatty, Paul. 1995. "Understanding the Standardized/ Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11: 147-160.
- 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., and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: SAGE Publications, Inc.
- Holstein, James A., and Jaber F. Gubrium. 1995. *The Active Interview*. Thousand Oaks, CA: SAGE Publications, Inc.
- Kvale, Steinar. 1994. "Ten Standard Objections to Qualitative Research Interviews." *Journal of Phenomenological Psychology* 25:147-173.
- Mishler, Elliot G. 1986. *Research Interviewing*. Cambridge, MA: Harvard University Press.
- 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. 367-391. New York: John Wiley.
- Schober, Michael F., and Frederick G. Conrad. In preparation. "Reducing Survey Measurement Error through Conversational Interaction."
- Suchman, Lucy, and Brigitte Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85 (409):232-253.

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.