

REDUCING RESPONSE EFFECTS IN TELEPHONE INTERVIEWS

Marianne Berry and Peter V. Miller
Survey Research Center, The University of Michigan

Telephone interviewing is fast becoming the norm, rather than the exception, in survey research. The motivations for choosing to do telephone surveys have been primarily their lower cost and greater ease of administration. The telephone generally has been regarded as a more cost-efficient alternative to the personal interview--a personal interview conducted at a distance, as it were.

However, the dynamics of interview interaction on the telephone are largely unexplored. Comparisons of telephone and personal interviewing, both those employing a verification approach and those comparing response differences obtained by the two modalities, generally show few differences.

Another body of literature relevant to this question considers the psychological consequences of the lack of visual contact. Investigation of nonverbal communication in naturalistic settings suggest that the visual channel is important in conveying affect and evaluation of others, and that visual cues affect impression formation and person perception.

But the implications of such findings for telephone interviewing are not clear-cut. The lack of visual cues may hinder the establishment of trust or rapport between the interviewer and the respondent; on the other hand, this depersonalization of the interview may result in less response bias by diminishing the importance of the interviewer's personal appearance. Moreover, if respondents find it easier to report on sensitive or embarrassing topics when the interviewer is not physically present, the telephone survey may actually prove to be a superior method for gathering this type of data.

The results presented here proceed from our first efforts to adapt to the telephone some techniques that have successfully improved the validity of responses given in personal interviews. These techniques were developed by Charles Cannell and associates of the Interviewing Methodology Program at the Survey Research Center. Three techniques were used in this study--commitment, instructions, and feedback--each of which is described briefly below. (For a more complete description of this program of research, see Cannell, C.F., L. Oksenberg, and J. Converse (eds.) Experiments in Interviewing Techniques. Ann Arbor: Survey Research Center, The University of Michigan, 1979.)

Commitment

A major source of response invalidity is respondents' inability or unwillingness to answer questions completely and accurately. The commitment procedure induces the respondent to make a conscious, overt act of agreement to carry out the activities that make up good respondent performance.

After the respondent selection procedure has been completed, before the beginning of the interview, the following statement was read:

In this research it's important for The University of Michigan to get accurate information--that is, to get exact details on every item, even on

those which may seem unimportant to you. It may take extra effort for you to report accurately. Are you willing to think carefully about each question in order to give accurate information?

Respondents were required to make some verbal signal of agreement before the interview could proceed. Before beginning the interview, we continued:

For our part, we will keep all information you give us very confidential. Of course, the interview is voluntary. Should we come to any question which you do not wish to answer, just let me know and we'll go on to the next one.

Instructions

We assume that at the beginning of the usual survey interview respondents have little knowledge of what is expected of them beyond the broad cultural norms about the interaction. Our use of detailed and frequent instructions had two objectives: to clarify the goals of the performance, and to clarify the specific tasks involved in achieving that goal. The general goals of the interview were first presented in the commitment statement, as above, and reiterated at various junctures throughout the interview. A second type of instruction gave more specific details as to how a respondent should go about answering a question, such as the following:

In answering the next few questions it may help you to be accurate if you consult a program listing or TV guide.

Feedback

The final technique employed was feedback. Much of the interaction that takes place in a survey interview is not governed by the strict asking and answering of questions. Reactions that respondents get from interviewers are important determinants of their subsequent behavior. Feedback statements were designed into the questionnaire, standardizing and controlling this important process. In general, feedback was made contingent on good performance, and both positive and negative feedbacks were employed. For example, respondents who took less than three seconds to answer questions which involved recall of distant events were read:

As I mentioned, sometimes it's hard for people to remember everything. Perhaps if you think about it a little more you will remember something you missed. Was there anything else?

Positive feedback, on the other hand, was used to indicate that the answer given fulfilled the goals of the question, as, for example:

"Thanks, this is just the sort of information we're looking for."

"I see. We appreciate your effort in remembering these details."

Previous research having shown that commitment, instructions, and feedback each improve aspects of

respondent performance, it seemed appropriate to assess their combined effects. However, given the absence of nonverbal cues on the phone, it was also important to examine the effects of feedback separately from commitment and instructions. These objectives mandated a design with two experimental groups, one featuring commitment and instructions alone, and the other adding feedback. We anticipated a linear improvement from control to partial experimental to full experimental group.

	<u>Commitment and Instructions</u>	<u>Feedback</u>
Control	--	--
Partial experi- mental	yes	--
Full experimental	yes	yes

Although this design does not permit us to separate the effects of commitment and of instructions separately, it does allow assessment of the effectiveness of the commitment and instructions alone, and of the increment in reporting accuracy created by the use of feedback.

Dependent variables

Previous research using these experimental techniques in personal interviews had been conducted using health variables for which validation data were available, and the effectiveness of the experimental manipulations was established by comparing information provided by respondents with some independent record of the same information. For many variables of interest, however, such as attitudinal items, such records do not exist. Another approach, the one used in this research, is to select dependent variables for which the direction of bias is well established in theory or previous research, but for which the magnitude of the invalidity is undetermined. Specifically, in addition to questions on the topic of health, we added a section about respondents' media habits. Respondents were asked how much they liked certain types of television programs, which were chosen for their positive or negative social desirability.

The Sample

The study used a computer-generated random digit dialing (RDD) design. The sample numbers assigned to each experimental group came from a common sample design. We constructed a clustered design of randomly generated telephone numbers from working central office codes in the coterminous United States, a design as described by Waksberg (1978). We used a cluster size of twelve, found by Groves (1979) to be desirable in the evaluation of design effects. To reduce sampling variance for differences between results in treatment groups, we randomly assigned four elements from each cluster of the three experimental groups.

Respondents were randomly selected from a household listing using the Kish selection procedure (1949), as adapted for the telephone. The response rate was 67%.

Data Collection

The data were collected in early spring of 1979 by Survey Research Center interviewers, using a computer-assisted telephone interviewing (CATI) system, in which the questionnaire is programmed to display the appropriate question, depending on the previous response. Interviewers

used a console to enter responses directly into a computer by typing either a numeric code or verbatim responses to open-ended questions.

ANALYSIS AND FINDINGS

To test the effects of the experimental interviewing procedures on reporting of various measures, we carried out regression analysis, with an ordinal dummy coding for the experimental treatments as independent variables. This procedure enables us to attach regression weights to the treatments *per se*. The experimental comparisons in this analysis, then, consist of commitment and instructions *vs.* the control procedure, and the added effect of feedback over and above the commitment and instructions treatment alone.¹

In addition to the analysis of experimental treatment effects, the data were also analyzed by sex of respondent, since this is the first study in which male respondents have been exposed to the experimental interviewing procedures. It was important for us to discover whether the experimental treatment effects are specified by sex.²

Table 1 presents mean experimental group differences in reporting of several health variables and alcohol consumption, derived from the regression analysis, for the total sample and for men and women separately.

Health Variables

The first two measures are termed "ambiguous" items, in that they place an implicit burden on the respondent to define categories and to decide whether his or her circumstances fit the definition. For example, the measure of ambiguous health conditions is an index of responses to closed questions, asking the respondent if he or she has experienced given ailments. The illnesses in these questions are somewhat vague in their description (e.g., "hay fever," "repeated sinus trouble," "repeated back trouble," "flu or influenza"), requiring respondents to clarify the terms in their own minds and decide whether they apply to their own health situation. We hypothesized that the experimental interviewing techniques would aid the respondent in this task, both by motivating them to expend the effort, and by communicating a broadly inclusive general definition of such terms. Therefore, we anticipated higher reporting levels for the "ambiguous" questions in the experimental groups.

The hypothesis for ambiguous health conditions receives support in the data, although the statistical significance of the difference between the control and experimental groups is marginal ($p < .10$). For the other item in this category, health symptoms, the findings give stronger support for the hypothesis. This measure summarizes the number of mentions to three questions asking for reports of things which "bothered" the respondent in various parts of the body during the two weeks preceding the interview. Once again, respondents are required to define such "things," and whether they "bothered" them sufficiently to report the ailments. We hypothesized that the communicative context provided by the experimental interviewing procedures would help the respondent perform this task, and consequently produce higher levels of reporting. The data for the whole sample indicate that the commitment and instructions procedures produced the expected effect ($p < .05$), although

the addition of feedback did not add to the effect.

It is interesting to note that for both of these variables there is also a "main effect" of sex on reporting. Women have higher reporting levels overall, indicating a lower threshold of reporting and/or the impact of real health. While the experimental comparisons seem to produce a more linear pattern (from control to commitment-instructions to feedback) for women on these indices, the sex-by-interviewing treatment interaction does not border on statistical significance for either variables.

The other health measures ("items requiring effort") present a different type of response task. The "effort" in this case involves more memory search, frankness, and simple cooperation than defining and clarification. For example, the first item--health conditions--summarized responses to several closed questions asking for reports of (relatively) well-defined ailments (viz., asthma, hemorrhoids, pneumonia, broken bones). To answer these questions, respondents must search their memories and be willing to report illnesses which may be embarrassing or threatening. For the "health behaviors" item, respondents were asked to list in an open question the actions they do (or don't) perform to take care of their health. We viewed this measure as one involving effort and cooperation, in that such actions may not be very salient for many respondents.

We hypothesized that the experimental interviewing procedures would help to motivate respondents to expend the effort necessary to report non-salient, or threatening items. Thus, we expected higher reporting levels for these two variables in the experimental conditions. For the sample as a whole, the hypothesis is supported for both measures ($p < .02$). Unlike the "ambiguous" items, there is no main effect of sex for these variables. Like these variables, there is no sex-by-experimental treatment interaction.

Alcohol Consumption

Under the general assumption that drinking behavior is likely to be underreported due to embarrassment or lack of salience, we expected the experimental interviewing treatments to produce higher reported levels of drinking different alcoholic beverages by motivating respondents to give more considered, honest answers. This general hypothesis is not supported, as seen in the Total Sample column of Table 1. There seems to be a tendency for men to evidence the hypothesized interviewing treatment effect for beer and liquor, but the effects for women appear curvilinear for the three experimental groups. In any event, these within-sex differences are not significant. There is a large main effect for sex in reporting both beer and liquor consumption ($p < .01$).

To summarize the analysis of Table 1, it appears that the experimental treatments are effective in changing reporting levels for several types of reporting tasks in health, but not for alcohol consumption. The addition of feedback to commitment and instructions does not appear to improve reporting. Finally, there does not appear to be any evidence that the effects of the interviewing procedures interact with the sex of the respondent.

Attitudes toward Television

Another hypothesized effect of the interviewing techniques is the reduction in response error due

to social desirability bias. This factor may be seen as the effect of perceived social norms on the truthfulness of reports of attitudes and behavior which go counter to those norms. Some respondents, sensitive to the fact that their ideas depart from what they view as appropriate or common ones, may distort their reports so as to conform with the perceived norm. This may lead them to exaggerate things which they think are positive, and understate negative items.

A test of ability of the interviewing techniques to reduce such tendencies involves establishing the likely direction of error and demonstrating that respondents exposed to the interviewing methods produce reports which, when compared to the control group, move the opposite way.

Attitudes toward television programs provide a good testing ground for this hypothesis. We asked respondents how much they liked to watch various types of programs, selected for their possible social desirability tendencies. To test the effect of the interviewing techniques on reporting of attitudes affected by presumed negative social desirability, we measured attitudinal responses to "police and crim," "sexy," and "game" shows. To examine the impact of the techniques on attitudes flavored by positive social desirability, we asked respondents about their liking for "news" and "educational and cultural" pro-grams. A final item, tapping attitudes toward "movies" was hypothesized to be unaffected by the techniques, since it seemed to have no clear social norm which might affect reporting.

The effect of sex on reporting is important to examine in this case since we expect men and women to have different perceived social norms for different items. We anticipated, for example, that women would express more positive attitudes toward "police and crime" and "sexy" shows in the experimental techniques on these items, then, would be to move the responses for females closer to those given by males. For "news" and "educational" shows, we anticipated less positive attitudes by both sexes in the experimental conditions.

The analysis for this test is a comparison of the proportion of respondents in each experimental group reporting that they like to watch a television program "very much" or "somewhat." The results are presented in Table 2. For two of the three items with an assumed negative social desirability bias, the experimental techniques produce more positive reports of attitudes. In the experimental groups, women respondents particularly report liking "police and crime" and "sexy" shows in greater numbers. The former effect is statistically significant ($p < .05$), while the other borders on significance.

As expected, the experimental technique had no effect on reporting of attitudes toward movies. Contrary to expectation, however, the technique also had no effect on reporting of attitudes toward game shows, or toward the two positively biased items, news and educational programs.

In summary, this first test of the effects of the interviewing techniques on attitude reporting gives some support for the idea that the techniques can improve reporting of items biased in a negative socially desirable direction. We need to perform other tests of this type to specify more carefully the parameters of the effects of these experimental techniques on this type of reporting task.

SUMMARY AND CONCLUSIONS

This study attempted several major innovations in the program of interviewing research undertaken at the Survey Research Center during the past several years. The first obvious departure from our past research was the application of the experimental interviewing techniques to a telephone interview. The dynamics of interview interaction in this mediated setting are still largely unknown. This attempt to test different structures of communication in the interview through combinations of commitment, instructions, and feedback must be viewed within this context. The fact that the feedback procedure did not improve reporting when combined with commitment and instructions in this study (as it did in earlier investigations) suggests the possibility that the mediated environment in a telephone interview demands a different style of communication from the face-to-face interview. Other explanations for the finding are also plausible, and we are attempting now to investigate some of them analytically. It is clear, however, that much more research on telephone interaction is required.

The second innovation in the program of interviewing research provided by this study is the application of the interviewing techniques in a heterogeneous sample of respondents. Our earlier studies have sought to pinpoint the effects of the techniques by researching a homogeneous pool of respondents (middle-class, white women), thus minimizing the impact of actual experience on reporting. It is heartening to find that the experimental procedures appear to have similar effects for other demographic groups, although there are hints of interactions between the techniques and sex of respondent which must be more closely examined.

Finally, this study sought to assess the effects of the interviewing techniques on reporting of a wider variety of variables than has been attempted in the past. In particular, our examination of attitude reporting gives some support for the hypothesis that the procedures can reduce response bias in such variables, if our assumptions about the direction of bias are correct. Other tests of this idea are required, especially since the experimental techniques seem to have been limited in this study to reducing negative social desirability bias in the attitude reporting. We also need to adopt new criteria for judging the success of the procedures in reducing response error when we consider attitude measures. It would be interesting to know, for example, whether these variables evidence more "construct" validity when measured with the experimental techniques. Testing this would involve examining the relationships among attitude measures which are hypothesized to relate to one another in a theory, within experimental groups interviewed with the experimental procedures.

Like most studies, this first investigation of experimental interviewing procedures in telephone interviews has offered some support for hypotheses and some puzzles. In subsequent analyses and future studies, we will seek to better understand the nature of effective interview interview interaction on the telephone.

TABLE 1
EFFECTS OF EXPERIMENTAL INTERVIEWING TREATMENT AND SEX ON REPORTING OF HEALTH MEASURES AND ALCOHOL CONSUMPTION (Mean scores)

Variable		Total sample	Men		Women	
<u>Health: "Ambiguous" items</u>						
Number of conditions	Control	1.023 (347)	.890	(145)	1.119	(202)
	: Comm. + Instructions	1.178 (325)	1.015	(133)	1.292	(192)
	Comm. + Inst. + Feedback	<u>1.211 (327)</u>	<u>.948</u>	<u>(145)</u>	<u>1.423</u>	<u>(182)</u>
All:		1.135 (999)	.948	(423)	1.273	(576)
Number of symptoms	Control	1.866 (351)	1.527	(148)	2.113	(203)
	: Comm. + Instructions	2.307 (322)	2.113	(133)	2.444	(189)
	Comm. + Inst. + Feedback	<u>2.256 (336)</u>	<u>1.832</u>	<u>(149)</u>	<u>2.594</u>	<u>(187)</u>
All:		2.137 (1009)	1.814	(430)	2.377	(579)
<u>Health: Items requiring effort</u>						
Number of conditions	Control	1.392 (347)	1.414	(145)	1.376	(202)
	: Comm. + Instructions	1.603 (325)	1.699	(133)	1.536	(192)
	Comm. + Inst. + Feedback	<u>1.560 (327)</u>	<u>1.600</u>	<u>(145)</u>	<u>1.527</u>	<u>(182)</u>
All:		1.516 (999)	1.567	(423)	1.477	(576)
Number of behaviors	Control	2.877 (350)	2.952	(147)	2.813	(203)
	: Comm. + Instructions	3.379 (322)	3.007	(134)	3.644	(188)
	Comm. + Ins. + Feedback	<u>3.479 (336)</u>	<u>3.463</u>	<u>(149)</u>	<u>3.492</u>	<u>(187)</u>
All:		3.238 (1008)	3.147	(430)	3.306	(587)

(table continued)

TABLE 1 - continued

Variable		Total sample	Men	Women
Alcohol consumption:				
Glasses of wine	Control	3.303 (350)	4.109 (147)	2.719 (203)
(in past month):	Comm. + Instructions	4.650 (323)	4.265 (132)	4.916 (191)
	Comm. + Inst. + Feedback	3.180 (333)	3.279 (147)	3.102 (186)
All:		3.695 (1006)	3.871 (426)	3.566 (580)
Glasses of beer	Control	7.530 (345)	12.063 (143)	4.322 (202)
(in past month):	Comm. + Instructions	7.118 (321)	14.267 (131)	2.189 (190)
	Comm. + Inst. + Feedback	7.473 (332)	14.088 (147)	2.216 (185)
All:		7.379 (998)	13.456 (421)	2.945 (577)
Ounces of	Control	4.741 (344)	7.683 (142)	2.673 (202)
liquor (in	Comm. + Instructions	5.780 (322)	7.121 (132)	4.847 (190)
month):	Comm. + Inst. + Feedback	5.889 (333)	8.651 (149)	3.652 (184)
All:		5.458 (999)	7.849 (423)	3.703 (576)

NOTE: Ns appear in parentheses.

TABLE 2
EFFECTS OF EXPERIMENTAL INTERVIEWING TREATMENT AND SEX ON REPORTING OF
ATTITUDES TOWARD TYPES OF TELEVISION PROGRAMS (percent
"liking" programs)

Variable	Total sample	Men	Women
"Police and Crime" programs:			
Control	52 (338)	65 (143)	43 (195)
Commitment + Instructions	63 (317)	68 (125)	59 (192)
Commitment + Instructions + Feedback	62 (328)	63 (145)	60 (183)
"Game" Shows:			
Control	63 (309)	57 (143)	66 (166)
Commitment + Instructions	60 (311)	53 (123)	66 (188)
Commitment + Instructions + Feedback	63 (325)	55 (144)	69 (181)
"Sexy" shows:			
Control	45 (334)	65 (139)	31 (193)
Commitment + Instructions	54 (312)	66 (124)	43 (188)
Commitment + Instructions + Feedback	56 (315)	69 (139)	43 (176)
Movies:			
Control	91 (332)	92 (143)	90 (189)
Commitment + Instructions + Feedback	90 (309)	89 (123)	90 (186)
Commitment + Instructions + Feedback	92 (320)	92 (143)	90 (177)
News:			
Control	92 (340)	89 (147)	92 (202)
Commitment + Instructions + Feedback	88 (317)	89 (129)	88 (188)
Commitment + Instructions + Feedback	92 (335)	94 (149)	90 (186)
Education and Cultural:			
Control	88 (343)	87 (144)	88 (199)
Commitment + Instructions + Feedback	90 (320)	86 (134)	93 (186)
Commitment + Instructions + Feedback	94 (327)	95 (146)	93 (181)

NOTE: Ns appear in parentheses

FOOTNOTES

¹The ordinal dummy coding was carried out as follows:

	D ₁	D ₂
Control	0	0
Commit. + Instr.	0	1
Commit. & Instr. + Feedback	1	1

The equation for assessing main effects of the interviewing procedures on a dependent variable is:

$$\gamma = \beta_0 + \beta_1 \text{Commitment \& Instr.} + \beta_2 \text{Feedbk.}$$

The mean scores for each experimental group can be derived from the equation:

$$\beta_0 = \text{Control mean.}$$

$$\beta_0 + \beta_1 = \text{Commitment + Instructions mean.}$$

$$\beta_0 + \beta_1 + \beta_2 = \text{Commit. + Instr. + Fdbk. mean.}$$

²The equation for the effects of the experimental interviewing techniques and sex of respondent is:

$$\gamma = \beta_0 + \beta_1 \text{Sex} + \beta_2 \text{(C+I)} + \beta_3 \text{Fdbk.} + \beta_4 \text{(Sex x (C+I))} + \beta_5 \text{(Sex x Feedback)}$$

The means for the groups are derived as follows:

	Males	Females
Control	β_0	$\beta_0 + \beta_1$
Commit. + Instr.	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_4$
(C+I) + Feedbk.	$\beta_0 + \beta_2 + \beta_3$	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$

REFERENCES

- Groves, R.M. "An Empirical Comparison of Two Telephone sample designs." *Journal of Marketing Research*, 15, 1978:622-31.
- Kish, L. "A Procedure for Objective Respondent Selection within the Household." *Journal of the American Statistical Association*, 44, 1949:380-87.
- Waksberg, J. "Sampling Methods for Random-Digit Dialing." *Journal of the American Statistical Association*, 73, 1978:40-46.