

THE ROLE OF RESPONDENT DEBRIEFING QUESTIONS IN QUESTIONNAIRE DEVELOPMENT

Jennifer Hess, U.S. Bureau of the Census and Eleanor Singer, University of Michigan¹
Jennifer Hess, U.S. Bureau of the Census, SRD/CSMR, Washington, DC 20233

Key Words: Questionnaire development, Debriefing questions, Reliability, Comprehension

INTRODUCTION

One-on-one cognitive interviewing techniques such as concurrent think-aloud interviews are increasingly being used in questionnaire development to assess respondents' comprehension of the terms used, their ability to retrieve the information requested, and the processes they use in doing so. The results of such cognitive interviews are then applied by the questionnaire developer to construct new questions or revise existing ones. Thus, cognitive techniques have been used primarily at the "front end" of questionnaire development, prior to the administration of the complete questionnaire in a pre- or field test. Although their usefulness in diagnosing problem questions has been widely recognized, they also have serious limitations. Because of the time needed to conduct and analyze such interviews, typically only a very small number of purposively chosen respondents is interviewed at the questionnaire development phase, and methods of analyzing such interviews and of drawing conclusions from them are not well codified.

Many years ago, Howard Schuman (1966) suggested a technique that he called the "random probe" to assess respondents' comprehension of closed-ended questions within the interview and to flesh out closed responses with qualitative detail. Each interviewer probed a set of randomly selected questions from the interview schedule. Interviewers were instructed in the use of standard nondirective probes for each of their assigned questions. The advantage of this technique is that comprehension (and other qualities) are assessed for each respondent to the survey, albeit for a small number of questions, and that every question in the survey is subjected to such an evaluation.

Toward a similar goal, Belson (1981) developed a "question testing" method. Specially trained interviewers visited respondents the day after they completed a questionnaire and conducted intensive interviews. These intensive interviews focused on seven questions selected for the test. Interviewers were trained to ferret out respondent "misunderstandings, omissions, distortions, etc." through the use of memory aids, detailed probing, and cross-checking of answers.

Much more recently, researchers have begun to use

"debriefing questions" following a standardized interview to assess respondents' reactions to the interview and to evaluate comprehension of survey questions and concepts (e.g., Esposito, et al, 1992). In this paper, we argue that debriefing questions can be used as a cognitive tool to assess the quality of the interview, and that because they can be administered to a large and representative sample, they are an important "back-end" supplement to the one-on-one interviews used in the early phases of questionnaire development (DeMaio and Rothgeb, 1993). This paper 1) reviews the literature to see how respondent debriefing questions have been used in other surveys; 2) presents results of a respondent debriefing conducted in conjunction with the pretest of the Food Security Supplement to the Current Population Survey, and 3) recommends that such debriefings be made a routine part of questionnaire development and evaluation, along with such techniques as behavior coding.

EXAMPLES OF PREVIOUS RESEARCH USING RESPONDENT DEBRIEFING

The debriefing conducted following the Census Bureau's 1992 Simplified Questionnaire Test (SQT) is an example of debriefing questions designed to measure respondents' reactions to a survey. The SQT compared mail return rates to several experimental versions of the 1990 census short form questionnaire, plus a control form. Telephone debriefing interviews were used to assess both respondents' and nonrespondents' reactions to the questionnaires (Bates, 1992). Those who responded to the survey were asked about their initial impressions of the envelope, whether any questions were difficult or sensitive, who completed the form, and whether they had a favorable or unfavorable impression of the form. Nonrespondents were asked whether the envelope was ever opened, whether the survey was ever started, whether any questions were difficult or sensitive, and why the household chose not to respond.

Debriefing questions have also been used to assess question sensitivity. Researchers were concerned that questions on paternity, which were suggested for inclusion in the Child Support Topical Module of the Survey of Income and Program Participation, were particularly sensitive. To assess respondent sensitivity, Miller and Davis (1994) included debriefing questions following the child support module. Respondents were

asked if there were any questions they felt uncomfortable answering and, if so, which questions they were and what it was about the question that made them uncomfortable. Additionally, respondents were asked if there were any questions they didn't want to answer and, if so, what it was about the questions they identified that made them not want to answer those questions.

Several researchers have included debriefing questions to measure respondent comprehension of selected terms or questions. As part of an evaluation of questionnaire pretest methods, Cannell and his colleagues (1989) used several methods, including debriefing questions, to test a series of health-related questions. Their debriefing questions, which they refer to as "special probes", included items to measure comprehension, information retrieval, and response category selection. Measures of comprehension included questions to measure the respondent's understanding of concepts such as "red meat", "general physical examination or checkup", or how respondents interpreted "an illness that kept you in bed for more than half of the day". Information retrieval probes were designed to assess how respondents arrived at their answers, whether they had difficulty answering the question, and how accurate they thought their answers were. Response category selection probes asked how difficult it was for respondents to choose among the response choices given.

Debriefing questions have also been used to assess how respondents interpret the scope of a survey. Researchers were concerned with the underreporting of crime victimization in the National Crime Survey (NCS). Debriefing questions and vignettes were included at the end of the original NCS and the experimental questionnaire. They measured how well each questionnaire conveyed the scope of the survey to respondents and evaluated which instrument facilitated respondent recall of incidents that respondents were likely to shield or forget (Martin et al., 1986).

Respondent debriefing may also be used to gain better understanding of concepts that may influence the broader design of a questionnaire, rather than a specific term or question. A series of vignettes were used during the early stages of the redesign of the Current Population Survey (CPS) to test respondents' understanding of the concept "work", especially with regard to marginal work activities. The vignettes were used to identify potential problems to address when redesigning the questionnaire (Martin, Campanelli, and Fay, 1991). They also were used later in the questionnaire development process to assess whether the redesigned questionnaire was better at capturing marginal work activities and excluding non-work

activities (Polivka and Martin, 1992).

The preceding examples are only a few of the ways in which respondent debriefing has been used to guide questionnaire design. In the following section, we describe how we used respondent debriefing questions to evaluate reliability and respondent comprehension of questions on the subjective experience of hunger in the United States.

BACKGROUND

The Food and Nutrition Service of the U.S. Department of Agriculture is sponsoring a new supplement to the CPS, known as the Food Security Supplement. The objective of the supplement is to measure household food security in the United States. Researchers at the USDA postulated several components of food insecurity and attempted to obtain measures of each of them in the supplement questionnaire, drawing on the work of other researchers in this area.

As part of the developmental work on this questionnaire, we conducted a pretest from August 27-31, 1994 on a purposive sample in six states: Florida, Mississippi, Tennessee, Texas, Arkansas, and New Mexico. Of the 596 sample cases, one-third were conducted in the field by Computer-Assisted Personal Interviewing (CAPI). The other two-thirds were conducted at the Census Bureau's Tucson Telephone Center by Computer-Assisted Telephone Interviewing (CATI). The overall response rate was 70.3 percent (83.8 percent for CAPI and 64.1 percent for CATI). (The large difference between CAPI and CATI response rates was due primarily to the non-working telephone numbers.)

Respondents were asked both the CPS labor force questions and the Food Security Supplement questions. Following the supplement, respondents were asked up to 17 debriefing questions. These were designed to measure (a) the reliability of respondents' answers to some of the survey questions, (b) respondent comprehension of selected survey questions and concepts, and (c) respondents' reactions to certain aspects of the questionnaire. (Examples of the latter are not included in this analysis.) Several of the debriefing questions were similar to probes that would typically be included in a think-aloud type of interview.

RESULTS

Reliability

The supplement contained three sets of questions that focus on the following behaviors: adults skipping meals or cutting the size of their meals, adults going a whole day without food, and children skipping meals. The questions asking about these behaviors followed the same pattern: 1) Did respondent (or other adults in the

household) do the behavior in the last 12 months? If yes, then 2) In how many months in the last 12 months did respondent (or other adults in the household) do the behavior, 3) Did respondent (or other adults in the household) do the behavior in the last 30 days? If yes, then 4) How many days in the last 30 days did they do the behavior? (This is a household level questionnaire in which respondents report for themselves as well as other members of the household, including children.)

We included several debriefing questions to assess the consistency with which these questions were answered. First, respondents who said they had skipped meals or cut the size of their meals in at least 2 of the last 12 months were asked the following debriefing question: "You said there were some months in the past 12 months when you (or other adults in your household) cut the size of your meals or skipped meals. Do you happen to remember which months these were?" Respondents who said they did remember were asked to indicate which months these were. We asked similar questions of respondents who reported that adults in the household had gone a whole day without food in at least 2 months in the past 12 months and those who reported that children in the household had skipped meals in at least 2 of the past 12 months.

Second, we were concerned that respondents might have varying frames of reference for the 12-month time period. We therefore included the following debriefing question to assess respondents' frame of reference: "We asked you several questions about things you might have done in the past 12 months like borrowing food, putting off paying a bill, or skipping or cutting the size of your meals. When you answered these questions, were you thinking of the 12 months starting September 1993 and ending August 1994, or were you thinking of the 12 months starting January 1993 and ending December 1993?"

Third, we were concerned that 30 days was too long a reference period for respondents to accurately report the number of times they had skipped or cut the size of their meals. We included the following open-ended question for respondents who said that adults in the household had skipped or cut the size of their meals at least once in the last 30 days: "We asked you how many times in the last 30 days you skipped or cut the size of meals. How did you figure out how many times you skipped or cut the size of meals in the last 30 days?"

Results from the debriefing indicated that respondents differed in their frame of reference and that the answers to the original survey question were not very reliable. Almost a quarter of the respondents failed to understand correctly the time period referred to by the questions asking about "the last 12 months" (N=295). Forty-seven

percent (N=45) of respondents could not remember the months during which they skipped or cut the size of meals; 53 percent (N=15) could not remember the months during which they went a whole day without food; and 20 percent (N=35) said they guessed or estimated the number of days in the last 30 days on which they had skipped or cut the size of their meals.

Comprehension

Several questions on the survey presented complex concepts that we thought might be problematic for respondents. To assess comprehension of these questions, we asked open-ended questions similar to Schuman's "random probes" and similar to those used in typical cognitive interviews. Interviewers were instructed to record answers verbatim.

In several cases we asked respondents to paraphrase questions. One example is, "We asked you whether you were ever hungry but you didn't eat because you couldn't afford enough food. Could you tell me in your own words what that question means to you?" Verbatim responses indicated that respondents did not really distinguish between "ever being hungry but not eating because they couldn't afford enough food" and "running out of money to buy food," which was a concept asked about in other survey questions. That is, they did not appear to attend to the feeling of hunger specified in the question, in all likelihood because they had to attend to too many concepts at the same time: being hungry, not eating, not eating because they could not afford food (and not for some other reason).

Another complex question we asked respondents to paraphrase was the following: "We asked you earlier if you ever ate the same thing for several days in a row because you only had a few different kinds of food on hand and didn't have money to get more. Could you tell me in your own words what that question means to you?" Almost half of the respondents (N=294) made no mention of lack of money as a reason for eating the same thing for several days in a row. Moreover, this question is intended to measure diminished diet quality due to consuming a limited variety of inexpensive foods such as rice, beans, and macaroni products. When we followed up to find out what it was that they ate for several days in a row, meat was included somewhere on the list of foods eaten by 37 percent (N=108) of those responding "yes" to the survey question. It is doubtful that this question was understood or answered as it was intended.

Another debriefing item included to measure comprehension was the following, "You told me earlier that you ran out of the foods that you needed to make a meal and you didn't have money to get more. Did you run out of food altogether, or did you have some

food but not the kinds needed to make a meal?" We wanted to know whether respondents discriminated between "running out of money to buy food" from "running out of money to make a meal." Responses indicated that respondents understood the survey question as intended. About 88 percent said they "had some food but not the kinds needed to make a meal". Almost 100 percent (N=98) of respondents mentioned meat as a food that was needed to make a meal; many of them gave this as the first mention in their response. In this regard, low-income families appear to resemble American families in general.

Examples of Revisions Made to the Survey Instrument

As a result of the respondent debriefing, we, in conjunction with the sponsors of the survey, made several changes to the questionnaire. Questions asking in how many months persons in the household had done a particular behavior were modified to take into account respondents' inability to give exact numbers. For example, if respondents reported having skipped or cut the size of meals in the past 12 months, rather than asking in how many months this happened, we now ask, "How often did this happen -- almost every month, some months but not every month, or in only 1 or 2 months?" A similar change was made to the questions about adults who said they'd gone a whole day without food in the past 12 months and children who were reported to have skipped meals in the past 12 months.

To measure how well this new question works, we included a debriefing item following the April 1995 supplement. The debriefing item reads, "You said that there were some months in the last 12 months when you cut the size of your meals or skipped meals because there wasn't enough money for food. Do you happen to remember in how many months this happened?" If respondents do remember, they are asked to give the number of months. We can then compare answers to the debriefing and survey questions to see to what extent there is agreement between them. This strategy should also permit a rough check on whether the new version of the question appears to introduce unwanted order effects into the responses. If, in response to the question asking for the precise number of months, fewer people say one month or two months than do so to the closed-ended question, the hypothesis of an order effect should at least be entertained and perhaps tested further.²

Another revision made because of respondent debriefing data was to the question on eating the same thing for several days in a row. The pretest question read, "In the last 12 months, since September 1993, did you ever eat the same thing for several days in a row because you only had a few different kinds of food on

hand and didn't have money to get more?" Because the debriefing questions on the field test indicated that respondents did not understand the intent of this question, we revised it as follows: "In the last 12 months; did you ever serve only a FEW KINDS of low-cost foods--like rice, beans, macaroni products, and bread or potatoes--for SEVERAL DAYS in a row because you couldn't afford anything else?"

To measure how well this new question works, we have also included an open-ended debriefing question on the April 1995 supplement. The debriefing question asks, "You said that in the last 12 months you sometimes served only a few kinds of food for several days in a row because you couldn't afford anything else. What kinds of food did you serve?" As on the pretest, the interviewers are instructed to record respondents' answers verbatim. Answers to the revised question can then be compared with answers to the original question to see whether the revision resulted in the intended improvement.

DISCUSSION

The literature shows that debriefing questions have been used to evaluate many aspects of surveys including respondent reactions to an entire questionnaire (e.g., Bates, 1992), sensitive questions within a topical module (e.g., Miller and Davis, 1994), broader survey concepts (e.g., Martin, Campanelli and Fay, 1991), and comprehension (e.g., Esposito et al., 1992). The literature has also documented some of the drawbacks with this methodology (e.g. Oksenberg, 1991). For example, only a limited number of questions can be probed to keep from unduly lengthening the questionnaire. This may become less of a problem as computerized survey technology advances, and researchers are able to randomly select a limited number of debriefing questions for each respondent. A second problem is that some types of debriefing questions require that the researcher have prior knowledge regarding which questions are likely to cause problems for respondents and what those problems are in order to design worthwhile debriefing questions (e.g., DeMaio and Rothgeb, 1993). A third problem is retrospection. For example, since the Simplified Questionnaire Test was a mailout/mailback survey, the debriefing questions were not administered immediately after the respondent had completed the survey. In some cases, several days had passed before the respondent was contacted for the follow-up interview, and it is questionable how much respondents actually remembered about receiving and filling out the questionnaire (Bates, 1992).

The problem of recall can also occur when debriefing questions are included at the end of the survey instrument. In the current study, we placed the

respondent debriefing questions after the supplement questions. The pretest served both operational and questionnaire evaluation purposes and we did not want to disrupt the flow of the questionnaire in order to evaluate the questions. Ideally, placing debriefing questions on reliability at the end of the survey and questions on comprehension immediately after the survey question of interest may be preferable from the standpoint of recall. However, the effect on subsequent questions of including debriefing questions in the survey instrument needs further study.

Another problem we confronted was small N's for selected items. The debriefing questions each respondent received depended on the respondent's answer to the survey question. Very few respondents were asked certain survey questions, which meant that few respondents were asked the corresponding debriefing questions. However, even our small N's were large in comparison to N's typical of cognitive interviews.

Another issue needing further study is the effect that debriefing questions have on respondents. Do respondents feel that these questions are "checking up on them" to see if they change their answers when asked a similar question the second time? Do debriefing questions make respondents uncomfortable in other ways?

In principle, the issues of evaluating questions are no different from other kinds of evaluations, but survey researchers have seldom confronted them head on. How reliable do answers have to be before the question is judged to be adequate? How many people have to understand a question before it should be included in a questionnaire? If two questions are getting at essentially the same information, should both of them be included, or not? Debriefing questions can help us to quantify reliability, comprehension, and redundancy, but they cannot answer the further questions raised by such measurement.

CONCLUSION

In addition to providing respondents' reactions to a survey, respondent debriefing questions are useful for assessing both reliability and comprehension. Open-ended questions employing standardized probes can provide valuable information to indicate whether questions and concepts are well understood. Debriefing can be used in a normal survey setting on a large and representative sample of respondents, and the answers can be coded and analyzed like any other item of survey data. For very little expense beyond that of the field test itself, respondent debriefing can be used to evaluate questionnaires undergoing revision or on completed surveys to provide an additional measure of response

quality. Thus, respondent debriefing questions provide a useful supplement to other quantitative measures of quality, such as behavior coding or item nonresponse. They are a logical extension of developmental pretesting activities, since only by including some measures like these in the final survey instrument is it possible to evaluate the effectiveness of the developmental activities themselves.

¹ This article reports results of research undertaken by staff members of the Census Bureau. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

² Schwarz, et al. (1992) have shown that self-administered questionnaires give rise to primacy effects--i.e., greater than expected endorsement of the first response alternative in a series--whereas telephone or face-to-face interviews give rise to recency effects--i.e., greater-than-expected endorsement of the last response alternative in a series. In the closed-ended version of this question, "one or two months" is the last response alternative offered to respondents on the telephone.

REFERENCES

Bates, Nancy, "The Simplified Questionnaire Test: Results from the Debriefing Interviews," internal report, Center for Survey Methods Research, U.S. Bureau of the Census, August 18, 1992.

Belson, William A., The Design and Understanding of Survey Questions, Gower Publishing Co. Ltd., Hants, England, 1981.

Cannell, Charles, et. al., "New Techniques for Pretesting Survey Questions," final report for the National Center for Health Services Research and Health Care Technology Assessment, August, 1989.

DeMaio, Theresa J., ed., Approaches to Developing Questionnaires, Statistical Policy Working Paper 10, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, November 1983.

DeMaio, Theresa J., and Jennifer M. Rothgeb, "Cognitive Interviewing Techniques: In the Lab and in the Field," paper presented at the Allerton Conference on Methods of Determining Cognitive Processes Used to Answer Questions, November 6-9, 1993.

Esposito, James, Jennifer Rothgeb, Anne E. Polivka,

Jennifer Hess, and Pamela Campanelli, "Methodologies for Evaluating Survey Questions: Some Lessons from the Redesign of the Current Population Survey," paper presented at the International Conference on Social Science Methodology, Trento, Italy, June 22-26, 1992.

Fowler, Floyd J. Jr., and Anthony M. Roman, "A Study of Approaches to Survey Question Evaluation," working paper for the U.S. Bureau of the Census, February 25, 1992.

Schwarz, Norbert, Hans-J. Hippler, and Elisabeth Noelle-Neumann, "A Cognitive Model of Response-Order Effects in Survey Measurement," in Context Effects in Social and Psychological Research, Norbert Schwarz and Seymour Sudman, eds., Springer-Verlag, New York, 1992.

Martin, Elizabeth, "Report on the Development of Alternative Screening Procedures for the National Crime Survey," internal report by the Bureau of Social Science Research for the National Crime Survey Redesign Program of the Department of Justice, Bureau of Justice Statistics, January 15, 1986.

Martin, Elizabeth A., Pamela C. Campanelli and Robert

E. Fay, "An Application of Rasch Analysis to Questionnaire Design: Using Vignettes to Study the Meaning of 'Work' in the Current Population Survey," The Statistician (1991) 40, pp 265-276.

Miller, Esther R., and Wendy L. Davis, "Findings from the Cognitive and Field Interview Research on Questions About 'Proof of Paternity'," internal memorandum, Center for Survey Methods Research, U.S. Bureau of the Census, March 17, 1994.

Oksenberg, Lois, Charles Cannell, and Graham Kalton, "New Strategies for Pretesting Survey Questions," Journal of Official Statistics, Vol. 7, No. 3, 1991, pp. 349-365.

Polivka, Anne E. and Elizabeth Martin, "The Use of Respondent Follow-Up Probes and Vignettes in Pretesting and Selecting Questions," paper presented at the 1992 annual meeting of the American Association for Public Opinion Research.

Schuman, Howard, "The Random Probe: A Technique for Evaluating the Validity of Closed Questions," American Sociological Review 31(2):218-222.