

Use of Laboratory Methods and Cognitive Science for the Design and Testing of Questionnaires

Handbook of Methods

Judith T. Lessler



R&D Report
Statistics Sweden
Research - Methods - Development
U/STM - 35

Från trycket Augusti 1987
Producent Statistiska centralbyrån, Enheten för statistiska metoder
Ansvarig utgivare Staffan Wahlström
Förfrågningar Gösta Forsman tel. 08 7834180

© 1987, Statistiska centralbyrån
ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1987

INLEDNING

TILL

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Föregångare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

Efterföljare:

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report, U/STM 1987:35. Use of laboratory methods and cognitive science for the design and testing of questionnaires, handbook of methods / Judith T. Lessler.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

USE OF LABORATORY METHODS AND COGNITIVE SCIENCE
FOR THE DESIGN AND TESTING OF QUESTIONNAIRES

HANDBOOK OF METHODS

Judith T. Lessler
June 1987

Prepared for
STATISTICS SWEDEN

USE OF LABORATORY METHODS AND COGNITIVE SCIENCE
FOR THE DESIGN AND TESTING OF QUESTIONNAIRES
HANDBOOK OF METHODS

1. Introduction

This paper discusses the use of laboratory methods and the techniques of cognitive science for the design and testing of survey questionnaires. A survey is a scientific measurement process, and the survey questionnaire along with the proscribed protocol for administering the questions is the key measurement instrument. In the social sciences it plays the same role that laboratory equipment, scales, electron microscopes, particle accelerators, mass spectrometers, and so on play in the natural sciences. Because of this key role, the quality of our surveys is dependent upon the quality of the measurement methods. And the use of surveys to gather the information that we need (either for policy decisions or for understanding) requires that we develop a measurement process that is precise and accurate.

Understanding and controlling the survey error associated with the measurement process requires:

- methods for developing and testing survey instruments;
- monitoring protocols (quality control methods) for detecting and rectifying problems during the data collection;
- and mathematical models and experimental designs that allow assessment of the magnitude of error in the measurements.

This paper focuses on the first component and discusses a new approach to the design and testing of questionnaires, namely the cognitive laboratory approach. This approach entails gaining a detailed understanding of the question-answering process and using this knowledge to both understand the limits of what can be learned by asking questions

and to develop better questions. The other two components--quality control and error assessment--remain important. They are not discussed; however, the cognitive laboratory approach could be used to develop methods for monitoring the process and quantifying the errors. For example, one might discover during a series of laboratory studies that short recall periods and self-response provide the best data. However, because of the increased cost of using such and a possible concomitant need to have large samples, one might not be able to afford using the better methods in the survey. However, some portion of the sample could be measured by the more accurate methods. The information from the more accurate methods could then be used in the estimation process to assess and adjust for errors in the overall survey. In this paper, however, I restrict the focus to choosing a single measurement method for the survey.

What is the cognitive laboratory approach? The term is used to imply a method of research rather than a setting. It is a new and evolving technique; however, its key features, as currently understood, are:

1. The attempt to gain a detailed understanding of the response task and the question-answering process. This is done using techniques from the cognitive sciences.
2. An experimental design approach to testing questionnaires in which research topics are based upon hypotheses about the response process and the generation of these hypotheses through previous observational studies or experimentation.
3. The use of small groups of respondents selected to span the range of characteristics known to influence the survey response and the random assignment of respondents to different treatments. Testing does not attempt to replicate the general conditions (as defined by Hansen, Hurwitz, and Berstad, 1961) of the survey.
4. Iterative testing techniques that quickly apply the results of previous rounds of experimentation. This is different from the field testing approach which, because of the complexity of administering a testing protocol that is dispersed in the field, has rather slow cycles.

5. The collection of validation information either through debriefings of respondents, reinterviews, or data from other sources and the use of this information and the experimental results in formal evaluation of alternate questionnaires.

This approach will have several advantages. It will contribute to the long term accumulation of knowledge by formalizing the testing mechanism. As we test similar hypotheses over a variety of subject matter areas and a variety of respondent subgroups, we will increase our understanding of the question-answering process. In addition, the laboratory approach has the potential for bringing the researcher and analyst into contact with the survey measurement instruments. Often survey questions are subject to little detailed testing and are based on invalid assumptions as to how people will respond to questions. To use a computer analogy, we often approach the respondent as if he has the information we need stored in his mind in an "on-line disk," sorted, classified, and scaled the way we need it for analysis. For example, a consumer expenditure survey may ask the respondent to report his total telephone expenses breaking them down into charges for basic services, equipment, local and long distance calls, and taxes. The person may be asked to report for some overall reference period such as the prior 3 months and also for each month in the 3 month period.

Such questions and others like it (How many times did you eat melon during the past 3 months? How much beer have you consumed during the past month? What is the total income of people living in this household considering all sources?) require detailed recall and subsequent calculations. The truth is that the respondent may not think of the events of his life in the units that we want to measure. The information may not be readily accessible in memory. He or she may not be able to attend to the task at the moment that the interviewer appears. And by asking questions in this manner we may be transferring

the burden of making the survey measurements onto the respondent. However, if we gain a detailed understanding of the comprehension, recall, and judgment that are required of survey respondents, we may be able to design questions that are easier to answer and which will result in more accurate measurements.

The remainder of the paper is organized as follows: There is a short background section on developments in the U. S. in this area. This is followed by a section on testing and revising existing questionnaires, a short section on developing questionnaires on new topics, and a section on practical methods. The paper is written as a handbook of methods rather than as a scientific discussion of the topics. As such it is based upon the results of several studies that I have been involved at the National Center for Health Statistics (NCHS) and the Research Triangle Institute (RTI). This discussion of methods supplements the material in the lectures that were presented at STATISTICS SWEDEN during June 1987.

2. Background

In 1983 and 1984 the U.S. Committee on National Statistics of the National Academy of Sciences held an Advanced Seminar on the Cognitive Aspects of Survey Methodology (CASM). This seminar brought together a group of cognitive scientists and survey methodologists to discuss potential linkages between the disciplines of survey research and cognitive psychology (Jabine, et al., 1984). A number of innovative ideas concerning collaboration between survey researchers and cognitive scientists were developed in the conference. Subsequent to the seminar, NCHS with funding from the National Science Foundation (NSF) undertook a number of studies that investigated the feasibility of using cognitive

laboratory methods for studying the survey measurement process. One of these studies (Lessler, Tourangeau, and Salter, 1986) investigated the use of the cognitive laboratory for the design and testing of questionnaires and demonstrated the usefulness of the method. Many of the methods discussed in this paper were initially developed in connection in that project and have subsequently been used at the Research Triangle Institute and other places for examining questionnaires.

During 1986, NCHS developed the National Laboratory for Collaborative Research in Cognition and Survey Measurement. The laboratory is jointly funded by NCHS and NSF. Located at NCHS headquarters in Hyattsville, Maryland, the National Laboratory sponsors a Collaborative Research Program which supports research on cognitive issues pertinent of the survey research process and a Questionnaire Design Research Laboratory which applies cognitive research methods to the design and testing of questionnaires. Similar research is also being carried out at several private research institutes including RTI.

The method is based upon obtaining a detailed understanding of the question-answering process. According to Sudman and Bradburn (1974) the question-answering process in a personal interview survey consists of two roles linked by a task--an interview role, a respondent role, and the task of transferring information from the respondent to the interviewer. Okensberg and Cannell (1977) present a model of the question answering process that is summarized in Exhibit 1. According to their model the question-answering process entails:

1. Comprehension of the question in which the respondent tries to understand what is being asked.
2. A cognitive processing step in which the respondent makes some decisions as to the information needed in order to give an accurate answer, attempts to recall information, and then formulates a response based upon the recalled information.

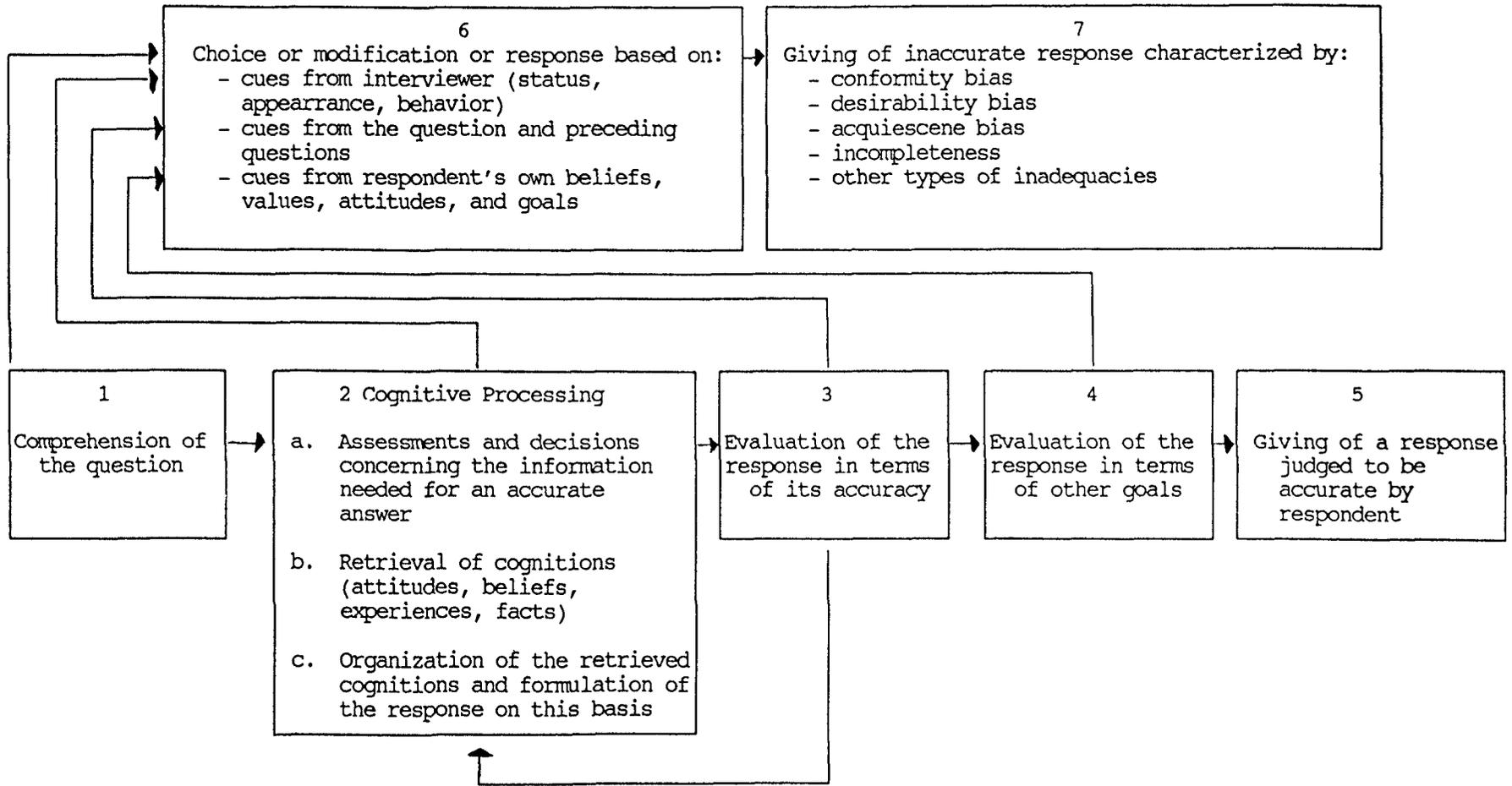


Figure 1. Model of the Question-Answering Process

3. An evaluation step in which the respondent judges the accuracy of the response. At this point the respondent may undertake additional thought--again assessing the type of information needed, attempting additional recall, and reformulating the response.
4. A second evaluation step in which the respondent may evaluate his response in terms of other goals. For example, the respondent may be hesitant to report certain events and attitudes.
5. The respondent may give a response that he believes to be accurate or may modify the response based upon other considerations.
6. Some of the things that may cause the respondent to modify his response include cues from the interviewer, context effects due to the structure of the questionnaire, the respondent's own beliefs and values, and so on.
7. The respondent may give a response that is inaccurate. The respondent may not be aware of how the things listed in step 6 produced a modification in the response.

This paper describes ways to obtain a detailed understanding of this process and methods for using that knowledge to develop better survey questions. We will focus on how one investigates the comprehension, recall, and judgment stages of the question-answering process and subsequently uses that information to design forms that are easier to answer.

The studies reviewed in the lectures demonstrated that if we find that respondents misunderstand certain terms, we can develop survey forms that more clearly define what is being asked. Or if we find that the length of the recall period makes it difficult for respondents to remember distinct events, we can shorten the recall period and provide cues to assist the respondent in the recall task. For example, it may be easy for a respondent to recall all of the doctor visits made by family members during a preceding 2-week reference period. However, he may have more difficulty if the reference period is 6-months or a year. We may be able to improve response for the longer recall period by

providing memory cues for the respondent such as asking about different kinds of doctor visits, the number of visits for general medical exams, the number for treatment of injuries, the number associated with chronic conditions, and so on.

The question answering process differs for different modes of data collection. Another benefit of understanding the process is that we gain insight into how problems with certain components of the process produce different types of survey errors. To illustrate this consider the model developed by Lessler and Holt (1987) for response to a mail questionnaire and reproduced in Exhibit 2.

In addition to the steps shown in the Okensberg and Cannell model, the successful respondent must complete a number of other tasks. He must read and understand both the questions and the instructions, understand the structure of the form, and must comprehend the question order. Often a questionnaire includes blocks of questions that apply to some persons and not to others. Thus, the respondent must interpret conditional statements or skip instructions in order to decide if a question applies. Lessler and Holt examined a draft questionnaire for the 1990 U.S. Census which includes a question that begins, "If you rent your house or apartment...." Respondents who are not renters are supposed to skip the question but are not explicitly instructed to do so. Some who should skip answer the question anyway. This need to find the appropriate questions occurs at several points in the response process and is shown in steps 3, 4, and 12 of Exhibit 2.

In a mail survey, as well as formulating a response, the respondent must write, mark, or code it on the questionnaire. This is shown in Step 10 of Exhibit 2. Respondents must also determine when all questions have been answered and must return the form.

What types of errors are associated with the components of this

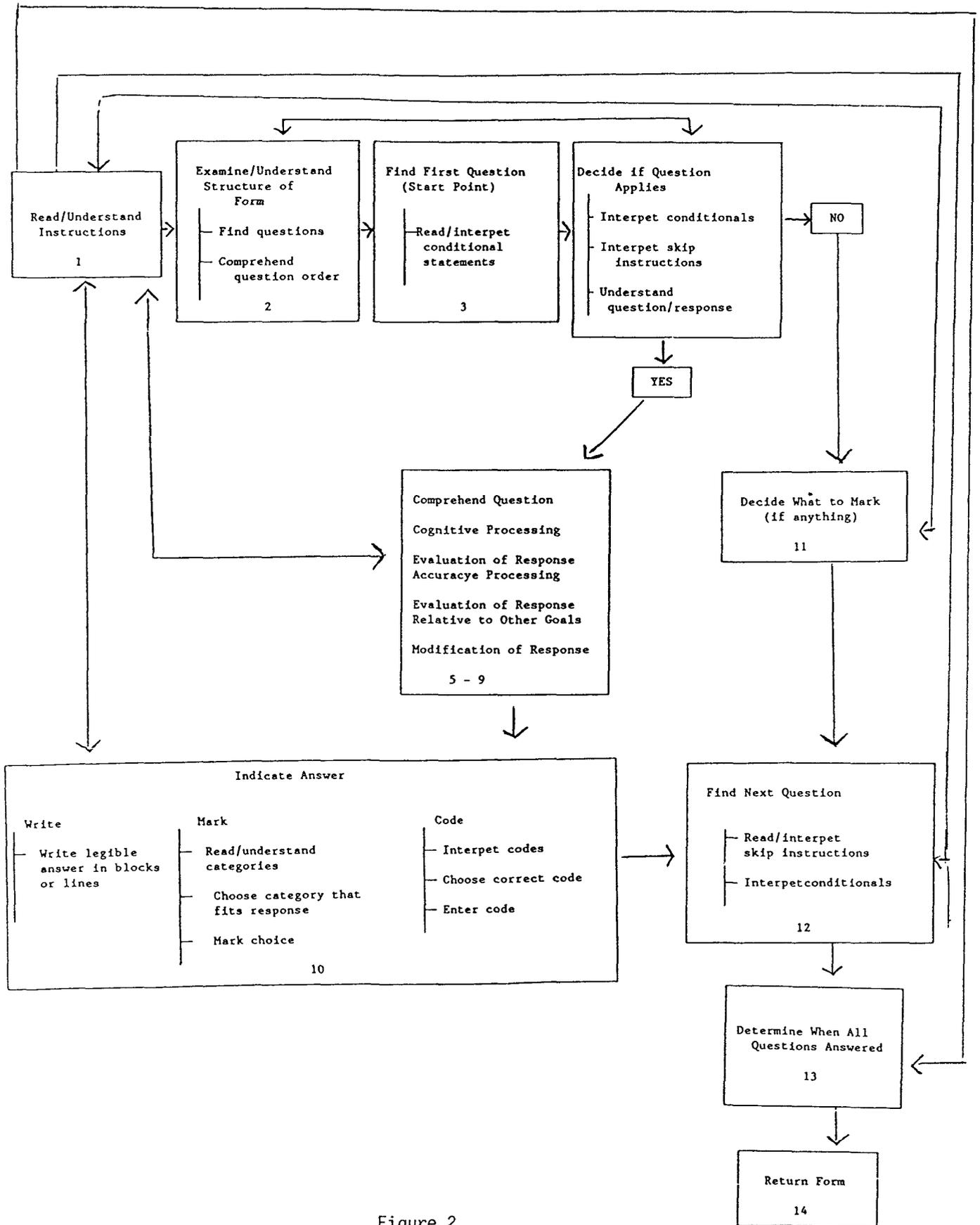


Figure 2

response process? Respondents who fail to understand the overall structure of the form will never complete it and will probably not return it. This will cause nonresponse errors. Failure to understand the "skip" instructions will cause the respondents to answer some blocks of questions that are inappropriate or to not answer some that they should answer. The former can be rectified in the during the editing of the data; the latter results in item nonresponse. Measurement errors can occur because of problems inherent in formulating a response and because of difficulties in indicating a response once it is formulated. Thus, models of this type can tell us where to focus our design efforts if we want to reduce a certain type of error. It may be that making the Census form easier to understand and begin will do more to increase the response rate than an extensive advertising campaign.

3. Evaluating an Existing Questionnaire

This section discusses the evaluation of an existing questionnaire, either a questionnaire that has been used for a number of years or a draft questionnaire for a new survey. I am going to make suggestions on how to do this evaluation and will not necessarily justify these suggestions. Some of the justification was presented in the lectures; others may be found in reports from various projects (Lessler, Tourangeau, and Salter, 1986; Tourangeau, Lessler, and Salter, 1986; Holt and Lessler, 1987).

3.1. Components of the Evaluation

During the evaluation we want to understand the process that respondents use when answering the questions. We want to identify

potential problems and delineate the good and bad strategies so that we can develop hypotheses about how to improve the process by changing the questioning strategy. I will call this the exploratory phase of the evaluation.

The next phase of the evaluation is devoted to experimental testing of alternative questions. Many lines of investigation will be identified during the exploratory phase, and it will be necessary to select specific ones for testing. The effort can be focused by classifying the problems and processes in terms of more general issues that are associated with several items in the questionnaire. For example, many items may require recall of the number of events in a reference period or may ask about things that some respondents are unfamiliar with. In this phase, many different questioning strategies are tested using volunteer respondents and experimental designs.

The final phase of the evaluation consists of a field test of a reduced set of alternate questions that are selected based upon the results of the experimental phase. The field test replicates the general conditions of the planned survey. It is also be experimental in nature but is distinguished from the prior round of investigations by the fact that the testing is conducted in the context of the general field conditions and focuses on fewer ways of making the measurements. For example, a telephone survey may be planned for the main survey but the initial rounds of testing may have involved personal interviews. In the final phase of testing, one would conduct the testing using telephone interviews.

Thus, we have identified three phases in the evaluation--an exploratory phase in which we make observations and develop hypotheses about the response process, a laboratory testing phase in which we use experimental design to test different questioning strategies in an

attempt to identify the most successful, and a field testing phase in which some of the ways of asking questions are tested under the conditions of the actual survey and which also uses experimental design. The following discusses each of these phase in turn.

3.2 Exploratory Studies

The goals of the exploratory studies are to understand the response process, identify any problems that respondents have with the response task, and to delineate good and bad strategies that are use by respondents to come up with answers. A good strategy is one that is likely to yield a response that meets the accuracy requirements of the survey. The exploratory studies are carried out by gathering detailed observations of the response process. Survey researchers have been collecting observations for many years. What distinguishes the process that I describe from previous methods is the explicit use of techniques from the cognitive sciences to increase the efficiency of the observational studies.

FORMS APPRAISAL: The first step in evaluating an existing questionnaire is for the research staff to carry out a forms appraisal. This consists of a formal review of the existing questionnaire during which its characteristics are described. The staff analyze the structure of the questionnaire including the process of moving from question to question and the structure of each question (open ended or closed). They note whether self or proxy response is required, and the tasks required by the respondents.

For example, consider the question "How many times have you been to the dentist during the past 12 months?" This is an open ended question that requires the respondent to report a number rather than choose a

response. In order to answer this question the respondent must understand what is included in the term "dentist." He or she must decide when 12 months ago was, not just in terms of a date but in terms of some event that he or she can then use to determine whether the recalled dental visits are in or out of the reference period. The respondent must then recall specific visits to the dentist, decide whether each is in the reference period, and count up the visits. This information must then be reported to the interviewer or, if the form is a self-administered form, recorded upon the form.

During the forms appraisal, list the types of problems that respondents might have and the types of mistakes that they might make. In the prior example, a respondent may have trouble recalling visits, may have difficulty anchoring the reference period, may recall visits but be unable to determine whether they are in the reference period, or may fail to recall some (or all) visits.

Classification is the next step in the forms appraisal. Classify each of the tasks required of the respondents into general types of tasks that are common to more than one question. Thus, you might note that several questions require determining the boundaries of a reference period; that others require the respondent to interpret a conditional statement, such as, "Answer this question if you operate a motor vehicle in connection with your job or business;" still others may ask the respondent to consider several types of events; others may require interpretations of unfamiliar terms; some may require calculations; some may require that respondents consult records; some may require opinions; others may be general knowledge questions (Who is the Prime Minister of Sweden?); and so on.

At the end of the forms appraisal you may have several things. You may have a flow chart that shows movement through the questionnaire.

You may have a question by question list of tasks, and also a classification of questions by type of required task. You will have notions as to what parts of the questionnaire may be difficult for respondents and what types of survey error could result if these difficulties are realized. You can use these initial ideas to focus the subsequent observational studies.

The amount of time that you can devote to the forms appraisal will probably be limited; therefore, you might wish to confine your study to certain subsets of the questions or to certain types of problems. For example, if you are evaluating a mail survey which has had high nonresponse on several items, you could focus your efforts on those specific items and attempt to determine the reasons for this nonresponse.

OBSERVING THE RESPONSE PROCESS: The next step in the evaluation is to observe some respondents completing the questionnaire. Since we are interested in what the respondents are thinking, we can not simply watch and listen to them. We must do something to get an insight into the respondent's thoughts.

Collection of response protocols or "think-aloud" interviews can be used to do this. In this technique, respondents are asked to report their thought processes as they answer questions. These are recorded either on audio or video tape and subsequently analyzed.

Respondents may initially have some difficulty in understanding what you mean when you ask them to think-aloud. They will need to be encouraged to do so, and it may be helpful to give an example. You can illustrate the technique by having the respondent work a simple math problem, by playing a tape of someone thinking aloud as they answer a question, or by giving an illustration yourself. If you give an illustration, do not use a question on the form that you are evaluating

because this may suggest a response process to the respondent. It is important to encourage the respondent to report his thoughts; however, you must be careful not to lead him.

Some questions will require little thought, for example, "How old are you?" Respondents will not report any thoughts when they answer this type of question since the answer is immediately accessible in memory. In other cases the respondent will need to think; however, the thought process is fast so do not let any time go by before you ask the respondent to report his thoughts. In the U.S., in a personal interview survey, respondents tend to look away when they start thinking, and one knows that "something is going on in their heads." If this is true in Sweden also, get them talking as soon as they look away or you will miss what they are thinking. A simple instruction said in a friendly way, such as, "Tell me what you are thinking?" should work.

People will, however, vary as to how well they are able to think-aloud and in their ability verbalize their thoughts. They may vary by their level of education but, from my experience so far, it seems that differences in educational levels cause differences in the type of language that is used rather than the amount of information given.

There are some people that have very little insight into what goes on in their minds. They will say things like, "I do not know why I say this. The answer was just there." Others tend to have visual images and will have trouble verbalizing their thoughts. However, most people seem to be able to do it when they get the idea as to what you want. You should probably choose one of your friends or acquaintances who is a "talker" for your first interview so that you can get practice before you take on the more difficult cases.

Except for notes as to what you want to ask about, do not attempt

to write down anything during the interview. Depend upon the recordings to document what is happening. There are two reasons for this. One, you are already introducing an artificial process by asking the person to think aloud which slows down (and may alter in other ways) the usual response process. If you are acting as an interviewer, you can write down the answers to the questions. However, if you take time to write down any of your observations about the response process (and there will be many), you will slow down the process even more. Second, you need to pay very close attention to what the respondent is doing so that you can ask questions about his thoughts and encourage him to think aloud. This brings us to the next technique that you want to use--detailed probes.

In some cases the respondent will not spontaneously report what you are interested in. For example, during the forms appraisal you may have decided that you are interested in how respondents anchor the reference period, and the respondent may not report this. Thus, you will need to ask questions about how he did this. You can ask something like "How did you know when a year ago was?"

You may find that the respondent did not do anything to decide when a year ago was. He just depended on a "general feeling." This is a bad strategy and not one that you want to encourage. Another respondent may say something like, "A year ago was June 5, and I was in France. I did not go to the dentist while I was there, so I just tried to remember visits since I came back from France." This is a good strategy since it provides a clear rule and a clear boundary for deciding what events are in the reference period. This is the type of strategy that you want to encourage in the questionnaire.

You may find that another respondent says, "June 5 was a year ago. I knew I made a visit to the dentist in summer or spring. I know that it is important not to miss any visits, so I reported it. It might have

been more than a year ago." Also a bad strategy, although not as bad as the first. And you have learned something else. This respondent thinks that the most important thing is to not miss any visits. Thus, he will be more likely to make errors of overreporting rather than underreporting.

Therefore, to supplement what the respondent mentions on his own, probe the respondent about his thought processes, his understanding of terms, his recall strategies, and his judgments. This will reveal difficulties that he may fail to mention during his spontaneous think alouds.

Collecting response protocols and doing detailed probing will take a long time--much longer than conducting an ordinary interview. Thus, you will not be able to collect response protocols and do detailed probing with each respondent for each question on the questionnaire. You can use other techniques to increase the number of observations. During the forms appraisal you may have developed some initial ideas as to how respondents approach the response task. You may think that they have difficulty understanding some terms or that they do not understand the skip patterns in the questionnaire. You can do some preliminary investigations of these hypotheses during the observational phase.

One way to do this is to is to conduct an ordinary interview and then use some post-interview technique to gain an understanding of the process. For example, you can ask the respondent to explain in his own words what was meant by certain questions or what types of information he thought the survey needed. If you think that the form contains a number of unfamiliar terms, you can ask about these terms directly or use a short comprehension test. This could be a short multiple choice quiz given at the end of the interview.

One can also investigate the response process by varying the

questions--either their wording or order. This can be done in the context of the collection of response protocols or in connection with a more abbreviated observational study. For example, in a self-administered form, you might remove all the skip instructions and see how respondents react. This will highlight the importance of using a clearly marked path through the questionnaire and give you better insight into problems that respondents have when they miss a particular skip instruction.

Who should you include in the initial observational studies? The respondents that are used for the exploratory studies need not be a random sample of the target population. However, it is important to select people who will vary on the characteristics that you are trying to measure. For example, if the survey asks about household expenditures, include people from large and small households, people with different incomes, people of different ages, and so on. Or, if you are asking about stress on the job, you need to include people of different occupations.

Start your initial observations with friends, co-workers, and acquaintances. Use these people to refine your observational techniques and gain experience in the technique. These people will be most helpful if they are willing to offer criticisms that you can use to improve your techniques. After that, use volunteers from the target population. Altogether some 10 to 50 respondents should be enough depending on the time you have and the number of things that you need to look at.

Analyze the observations by identifying the types of processes used, the strengths and weaknesses of various approaches, and the problems encountered by respondents. Since you have made observations on few respondents that are not necessarily a random sample of the

target population, focus on classifying and understanding the types of behaviors and thoughts and not on determining the number or proportion of people who experience particular problems. You of course will get some idea of frequency of problems--no one could answer Question 12--however, all you really know at this point is what is occurring not how often it occurs.

What you do is review the recordings noting the process used and the problems encountered. If you can get the response protocols transcribed, this will increase the efficiency of the analysis. As you review the observations, you may not notice something until maybe the sixth tape that you listen to, and will then want to go back and determine if other respondents also used a certain process. This will be easier if you can look at a transcription of the response protocols. The actual recordings will also be important because, as well as the words, the tone of voice, the pauses, and (if you have video) the actions of the respondents will be valuable in understanding what is going on.

Use the results to generate hypotheses as to how different question wordings, structures, and so on might improve the response process. These will be tested in the next phase. For example, you might think that you can improve response by changing the order of questions so that there are fewer skip patterns. If it appears that most respondents indiscriminately report every single event they recall so that the survey is beset with overreporting, you could test some questions that you hypothesize will cause respondents to be more discriminating in their decisions as to what to include in their response.

3.3. Experimental Testing

The goal of the experimental testing phase of the questionnaire evaluation is to test the hypotheses that you previously generated as to the more effective questioning strategies. You can also use experiments to further examine the response process by translating your hypotheses about the response process into different types of questions that can be compared. The different questions (or different measurement process) are then your experimental treatments. Recruit a group of respondents and randomly assign them to the treatments.

You will need a means of evaluating the results. For example, if your treatment is designed to reduce item nonresponse, the item response rate under the different treatments is the outcome variable. In other cases, you might collect validation information. For example, if you are interested in the reporting of dental visits, you can determine the actual number of dental visits from health care records or records from insurance payments. You may have information from other studies that indicates that under- or overreporting is the prevalent problem with the usual form of the questionnaire. Then the preferred method is the one which, respectively, produced lower or higher reports. You can also use a post-interview debriefings to evaluate the quality of the responses. For example, if you had learned through the observations that respondents had trouble understanding certain terms you can try different ways of defining the terms and, after the interview, administer a comprehension test.

In other cases, you may not be interested in determining whether one or another question is giving the "best" response, but rather you may be interested in determining whether you can produce a certain type of effect. For example, can you increase reporting by repeatedly encouraging the respondent to report events. Or you may be wondering if respondents are skipping questions at the end of a long

self-administered questionnaire because they are tired; so, you may decide to vary the order of the questions.

When I speak of a laboratory setting, I mean some setting that facilitates testing of different versions and not necessarily a place. Thus, you could ask respondents to come to STATISTICS SWEDEN for the interviews; you could conduct the interviews in a room in the respondent's neighborhood or community; or you might want to actually go into the households if you believe the response process is especially sensitive to the setting in which the interview is conducted.

Since you will probably have several versions of each question that you want to test, there will be many different questionnaires that you could construct. Also, you may want to change the questions after one or more rounds of experimentation. One thing that can be useful in developing several different versions of the questionnaire is to use a microcomputer. In the study reported by Tourangeau, Lessler, and Salter (1985), we developed 48 different versions of a dental health questionnaire by creating a computer "spreadsheet" file in which we had a single list that contained all versions of each question and a matrix that indicated the combination of question options that was to be included in the a particular version of the questionnaire. A print program then read a line of the matrix, selected the appropriate question options, and printed a version of the questionnaire. This prevented errors in treatment assignment that would have occurred if interviewers had been required to assemble or collate a questionnaire.

Make your tests realistic by keeping in mind the field setting for the main survey. Otherwise, you may identify a questioning strategy that works very well in the laboratory setting but does not work in the field. To give an extreme example, if you were developing a questionnaire for a telephone survey you would not want to test

different types of visual aids. No one would ever make this mistake; however, I give the example to remind you that there are circumstances associated with the mode, setting, and conditions of the interview that you will need to keep in mind when designing the experiments.

If you are testing a survey that has been ongoing for some time, one way to keep in touch with actual interviewing environment is to have some of the field staff comment on the different techniques that you are considering. Another way is to observe the field staff as they conduct a few interviews in the field setting using some of the questions that you are thinking of testing.

In the design and analysis of your study you will face problems of statistical power. Some of the differences between the treatments that you are using may not be large enough to detect with sample sizes that are available. You can relax your standards for detecting differences and use an alpha level of 0.15 or 0.20 instead of 0.05 or 0.10. However, if one method of collecting the data is much more costly than another, you might want to be more cautious in declaring it the best because, with a limited budget, using a more expensive measurement method will reduce the affordability of large samples. Otherwise, there is not much risk in using lower alpha levels when performing the statistical tests.

The laboratory experimentation is probably most useful if iterative rounds of testing are carried out and if observational techniques are also used. Thus, you may want to use observers, recordings, and debriefing of the respondents to supplement the experimental testing. This will allow you to generate additional hypotheses as to the nature of the response process and how it has been influenced by the different questions that you are using. Thus, if you could afford to test 300 respondents it might be better to do 2 successive experiments with 150

in each.

The sample sizes that you will need will of course depend upon the type and sizes of the effects that you are trying to detect. In every case, you will need to make sure that the range of respondents used in the experiments spans the range of behaviors that you are trying to study.

3.4. Field Testing

The last phase in the evaluation process is a field test. A field tests is very important if you are beginning a new survey or are making major changes in an existing one. The goal of the field test is to examine alternate methods under conditions that approach the general conditions of the actual survey. Thus one uses the type of staff that will be used in the actual survey, trains them in a manner similar to that which will be employed in the actual survey, selects people from the target population (although you may use purposively selected rather than random samples), and codes and tabulates the data as they will be coded and tabulated in the actual survey.

As well as using the general conditions of the survey, the field test employs fewer alternatives than the laboratory testing. This is because of the problems of conducting an experiment with a large number of treatments over a widely dispersed area.

Often in the U.S. pretests are done with a single version of a questionnaire. I think that this should never be done, especially when a new survey is being fielded. If the set of methods that are tested do not work, there is no alternate set of tested methods that can be used as a basis for comparison. You may not have two entirely different questionnaires, but certainly you should look at alternate questions for

the items that you are most concerned about. Presumably, if you have been through the other rounds of evaluation, you have more than one idea as to what will improve response. Included these in the field test--otherwise if things do not turn out as you expect, you may be left with little evidence as to why. For example, suppose you are looking for a way to reduce item nonresponse in a survey that has been conducted in the past. You may have concluded from the laboratory testing that certain type of questioning strategy will prevent or reduce the level of item nonresponse. When you test this in the field, you could find that item nonresponse is still high and conclude that your laboratory testing failed to identify a good technique for reducing item nonresponse. This may not be the case at all. Rather it may be that conditions have changed in the general public, that the overall level of nonresponse has risen, and that your new method is still to be preferred over the old method.

Thus, the field test should also employ an experimental design with subsamples being randomly assigned to the different measurement methods (different versions of the questionnaire). Again, the required sample sizes will depend upon the sizes and types of effects under study.

3.5. Summary

In summary, this approach to the development of and testing of questionnaires includes:

1. Observational studies in which the details of the response process are investigated and problems identified.
2. An experimental phase in which hypotheses about how the response process can be improved are tested using an experimental design. Alternate sets of questions are selected for the field testing phase.
3. A field testing phase that examines these alternatives. A

preferred questionnaire is then selected for the survey.

This is the ideal testing program. Many small scale surveys can not afford such an extensive testing program. However, even small scale development efforts will benefit from the use of the observational techniques described above. Sometimes collecting response protocols with only a few people can identify problems with the questionnaire and suggest solutions. Experiments on subsets of the items or key variables can be used if the budget is limited. The point is, as far as possible, questionnaire development activities should be based upon actual observation and experimentation.

4. Developing New Questionnaires

For a new survey there is no draft questionnaire available. One must be developed. Basic decisions must be made including specification of the target population, the mode of interviewing, the data collection period, and so on. Probably the first thing a survey methodologist charged with the design of a questionnaire does is to search for questionnaires on the same or a similar topic. Other techniques for developing a new questionnaire are unstructured interviews and focus groups. In each of these the researcher talks to members of the target population and examines how they think about the survey topics, the types of language used, their concerns, and so on. There is a good discussion of these methods in "Approaches to the Design of Questionnaires" published by the U.S. Office of Management and Budget.

Unstructured interviews are usually conducted with a guide for the topics that the researcher wants to cover. The respondent is encouraged to discuss the topics in his own terms. This allows one to note the language that is naturally used by the respondents and to get an idea of

how the respondents think about the survey topics, the way their memories are organized, and the salient features of the topic. It is generally believed that recall will be better if questionnaires are organized the way memories are organized. Things that are more important to the respondent are easier to recall, and the questionnaire will need to provide fewer cues for these topics. Again, it is best to record the unstructured interviews so that they can be analyzed later.

Focus groups are used to gather the opinions and ideas of a group of people. The groups should include from 6 to 10 people and be somewhat homogeneous as to type of person. The reason for this is that if there is a wide range of age and educational levels in a group, the older more educated people will tend to dominate the conversation. At RTI, we often have two moderators who use a discussion guide to direct the conversation. We record the sessions and ask that only one person speak at a time.

The results of these conversations with members of the target population and the needs of the analysts can then be used to develop initial drafts of the questionnaire which can then be tested using the three phases described above.

5. Practical Matters

If you want to start using some of these techniques you may want to begin by reading some background material. A list of references is included in the appendix. The report by Jabine et al.(1984) reviews the results of the CASM conference and contains some papers on the potential benefits of using the methods of cognitive science for studying survey measurement.

You should not be hesitant to begin to apply this technique because

you are not experienced. It is very new, and there are few people that are experienced. Conducting the think-aloud interviews and the detailed probing will probably be the newest thing that you try. You can learn how to do this by reviewing your own tapes. Start with some interviews of colleagues. Review the tapes looking for what you did that encouraged the respondents to report their thoughts and things that you did that were not effective. After you have done a few interviews, it is easy to begin leading the respondent. You have heard what other people have said, and when, a particular respondent is searching for words to try and express his thoughts you may be tempted to suggest something to him. The respondent may accept your suggestion when it was not what he had in mind at all. Another good way to learn is to work in pairs or groups and evaluate each other. Also, being a respondent in a think aloud interview provides good insight into the process.

Cognitive psychologists who have studied memory, comprehension, and decision making will be helpful as consultants. If you can find someone who has studied autobiographical memory, that person will be particularly helpful. Regardless of their specialty, cognitive psychologists can provide good advice in experimental methods. I will try to find out who might be useful to contact in Sweden.

There are several ways to get respondents. You could start with employees of STATISTICS SWEDEN since they are readily accessible. This is good group to use to develop your methods before contacting the general public. In all of the projects that I have worked on we have paid respondents who come from the general public. We have recruited respondents by placing ads in newspapers, putting up posters in public places advertising the need for volunteers, calling a sample of telephone numbers and recruiting a quota sample, and contacting clubs, community colleges, church and civic groups. In the last case the

organization is given the payments rather than the individuals. We try to identify people who cover the range of behaviors that we are trying to measure in the survey and who span the range of characteristics that are known to influence response in surveys.

As you start contacting people you may find some who are willing to participate but who do not have the characteristics that you need for the current study. You might want to keep a list of these potential respondents. These people could then be called and invited to participate in a later study.

At RTI we use telephone interviewers to screen the potential respondents. We train them to clearly explain what will be expected of the respondent but not to reveal the exact topic of the survey since we do not want the respondents to be thinking about the topic ahead of time.

Some of the people who make appointments for interviews will not come. We generally send letters and call respondents to remind them of their appointments. This also helps to identify people who have changed their minds and decided not to attend the session.

We generally pay people at the end of the interview--either in cash or by check.

The think-aloud interviews require privacy and recording equipment. Otherwise they can be conducted almost anywhere--STATISTICS SWEDEN or in some other place convenient to the respondent. I have recorded interviews using both video and audio tape. Some people take a while to become comfortable with the video tape. We obtain written permission for playing the tapes for other people and promise to only use the tape for statistical research purposes. We usually ask for permission to record the session at the beginning and wait until after the session to request the permission to share the tape with other researchers. This

is because a respondent is not willing to give permission to share the recording until he knows what it will contain. We also offer to let the respondent listen or watch his own tape before giving the permission. In my experience, the only people who have wanted to do so are those who have never seen themselves on video.

I have had one person refuse to give permission to share his answers. This person was concerned that he did not make a "good impression," because he could not recall some of the information requested. This brings up another important point. It is VERY important to convey to the respondent that you are evaluating the questionnaire not the respondent. The respondent should be told this from the beginning, and the first time he says something that indicates that he is having some difficulty recalling information or understanding what he is supposed to do, you should praise him saying something like, "That is good information. That is exactly the kind of thing that we want to hear about." Or "Good. That's just what we need to know." This lets the respondent know that the more he tells you about his difficulties the better. Some will still be embarrassed that they cannot do the task, and you will have to work to decrease their embarrassment.

Finally, you may want to make some effort to get the interviewing staff involved in the evaluation process, particularly if you are contemplating changes in an ongoing survey. Sometimes the field staff will resist changes because they do not understand the reasons for the changes and because it means more work for them. Thus, if experienced interviewers or field supervisors participate in the observational and experimental phases of the evaluation, they will be more willing participants during the field testing phase.

Have fun!

REFERENCES

- Jabine, Thomas B., Miron Straf, Judith M. Tanur, and Roger Tourangeau (1984) Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, National Academy Press: Washington, D.C.
Available from: Committee on National Statistics
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C 20418
- DeMaio, Theresa J. ed.(1983) Approaches to Developing Questions, Statistical Policy Working Paper 10, Statistical Policy Office, Office of Management and Budget, Washington, D.C.
Available from: Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402
- Lessler, Judith T. and Mimi Holt (1987) "Using Response Protocols to Identify Problems in the U.S. Census Longform," paper prepared for the Survey Research Methods Section, American Statistical Association Meeting, San Francisco. Will appear in the 1987 Proceedings.
- Tourangeau, Roger, Judith T. Lessler and William J. Salter (1985) Cognitive Aspects of Questionnaire Design: Part B Report, National Opinion Research Center, Chicago. Paper prepared for the National Center for Health Statistics under Contract No. 282-84-2123.
- Okensberg, Lois and Charles Cannell (1977) "Some Factors Underlying the Validity of Response in Self Report," Bulletin of the International Statistical Institute, pp. 325-346.
- Lessler, Judith T., Roger Tourangeau, and William J. Salter (1985) "Control of Errors in Surveys," Proceedings of the 1986 Conference on Survey Research Methods in Agriculture.
- Hansen, M., W.N.Hurwitz, and M.A. Bershad (1961) "Measurement Errors in Census and Surveys," Bulletin of the International Statistical Institute, Vol,38, pp. 359-374.
- Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell (1987) "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys," Science, April 10, 1987 pp.157-161.
- There are 4 articles on the cognitive aspects of survey methodology in the Milbank Memorial Fund Quarterly: Health and Society, Volume 63, Number 3, Summer 1985.
- Loftus, Elizabeth F. (1982) "Memory and its Distortions" in Kruat, A.G.(Ed.), The Stanley Hall Lecture Series, American Psychological Association, Washington, D.C.
- Lessler, Judith T. and Richard A. Kulka (1983) "Reducing the Cost of Studying Measurement Error: Is a Laboratory Approach the Answer?" in Wright, Thomas (Ed.), Statistical Methods and the Improvement of Data Quality, Academic Press, Inc.

Tidigare nummer av Promemorior från U/STM:

NR

- 1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)
- 2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)
- 3 En presentation av Box-Jenkins metod för analys och prognos av tidsserier. Åke Holmén (1979-12-20)
- 4 Handledning i AID-analys. Anders Norberg (1980-10-22)
- 5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)
- 6 Metoder för evalvering av noggrannheten i SCBs statistik. En översikt. Jörgen Dalén (1981-03-02)
- 7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)
- 8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)
- 9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan Wretman och Göran Örnberg (1983-06-20)
- 10 Current survey research at Statistics Sweden. Lars Lyberg, Bengt Swensson och Jan Håkan Wretman (1983-09-01)
- 11 Utjämningsmetoder vid nivåkorrigering av tidsserier med tillämpning på nationalräkenskapsdata. Lars-Otto Sjöberg (1984-01-11)
- 12 Regressionsanalys för f d statistikstuderande. Harry Lütjohann (1984-02-01)
- 13 Estimating Gini and Entropy inequality parameters. Fredrik Nygård och Arne Sandström (1985-01-09)
- 14 Income inequality measures based on sample surveys. Fredrik Nygård och Arne Sandström (1985-05-20)
- 15 Granskning och evalvering av surveymodeller, tiden före 1960. Gösta Forsman (1985-05-30)
- 16 Variance estimators of the Gini coefficient - simple random sampling. Arne Sandström, Jan Wretman och Bertil Waldén (Memo, Februari 1985)
- 17 Variance estimators of the Gini coefficient - probability sampling. Arne Sandström, Jan Wretman och Bertil Waldén (1985-07-05)
- 18 Reconciling tables and margins using least-squares. Harry Lütjohann (1985-08-01)

- 19 Ersättningsens och uppgiftslämnarbördans betydelse för kvaliteten i undersökningarna om hushållens utgifter. Håkan L. Lindström (1985-11-29)
- 20 A general view of estimation for two phases of selection. Carl-Erik Särndal och Bengt Swensson (1985-12-05)
- 21 On the use of automated coding at Statistics Sweden. Lars Lyberg (1986-01-16)
- 22 Quality Control of Coding Operations at Statistics Sweden. Lars Lyberg (1986-03-20)
- 23 A General View of Nonresponse Bias in Some Sample Surveys of the Swedish Population. Håkan L Lindström (1986-05-16)
- 24 Nonresponse rates in 1970 - 1985 in surveys of Individuals and Households. Håkan L. Lindström och Pat Dean (1986-06-06)
- 25 Two Evaluation Studies of Small Area Estimation Methods: The Case of Estimating Population Characteristics in Swedish Municipalities for the Intercensal Period. Sixten Lundström (1986-10-14)
- 26 A Survey Practitioner's Notion of Nonresponse. Richard Platek (1986-10-20)
- 27 Factors to be Considered in Developing a Reinterview Program and Interviewer Debriefings at SCB. Dawn D. Nelson (1986-10-24)
- 28 Utredning kring bortfallet i samband med den s k Metropolitdebatten. Sixten Lundström (1986-11-10)
- 29 Om HINK-undersökningens design- och allokeringsproblematik. Bengt Rosén (1986-10-06)
- 30 Om HINK-undersökningens estimationsförfarande. Bengt Rosén (1987-01-28)
- 31 Program för minskat bortfall i SCBs individ- och hushållsundersökningar - förslag från aktionsgruppen för uppgiftslämnarfrågor. Lars Lundgren (1987-02-20)
- 32 Practical estimators of a population total which reduce the impact of large observations. Jörgen Dalén (87-03-03)
- 33 A Multivariate Approach to Social Inequality. M Ribe (1987-05-05).
- 34 BORTFALLSBAROMETERN NR 2. Bortfall vid I-avdelningen, AKU och F-avdelningen t o m 1986. Håkan L. Lindström, Peter Lundquist och Tomas Garås (1987-08-01)

Kvarvarande exemplar av ovanstående promemorior kan rekvideras från
Elseliv Lindfors, U/STM, SCB, 115 81 Stockholm, eller per telefon
08 7834178