

INFORMAL TESTING

Dawn D. Nelson, Bureau of the Census

Considerable concern has been expressed over the inability to quantify the effects of non-sampling error on survey estimates. The potential sources of nonsampling error are known, however, and can be addressed by methods such as field testing before the survey is conducted. For example, field testing can be used to eliminate questionnaire problems that might cause response errors. However, in the real world of budgets and deadlines, there may be a strong temptation to ignore this step in developing the questionnaire. Unfortunately, the trade-off may be lower quality or worthless survey data. Thus, the time and money spent on field testing should be viewed as cost-effective in the long run. Also, the testing does not necessarily need to be complex or expensive to produce useful results.

A relatively simple type of testing is discussed in this paper and is referred to here as "informal testing." The definition of an informal test used in this paper is: a questionnaire field test involving a relatively small number of interviews in the kind of setting chosen for the final survey (i.e., home, work, etc.) as opposed to a laboratory setting. In this type of testing, the detection and correction of problems in the questionnaire draft depends mainly upon subjective information provided by interviewers or observers. The test is not designed to be evaluated on a rigorous statistical basis.

Informal testing is appropriate for use in the development of face-to-face and telephone questionnaires. It is not as useful in developing mail questionnaires because the information necessary for evaluating the questionnaire cannot be obtained from interviewers and observers.

The purpose of this paper is to encourage the use of this type of questionnaire testing when more sophisticated testing, such as split-sample testing, is too expensive or is not necessary to meet the test objectives. Section 1 of the paper describes the preparations before an informal test; Section 2 covers the operation and evaluation of the test; Section 3 discusses time, cost, and staffing considerations; and Section 4 provides an example of an informal test conducted in preparation for a recent nationwide survey.

1. PREPARING FOR AN INFORMAL TEST

Identifying the Objectives of the Test

Informal tests are particularly appropriate and useful in discovering poor question wording or ordering, errors in questionnaire layout or instructions, and negative response effects caused by the length of the interview or a respondent's inability or unwillingness to answer the questions. In addition, they can be used to a lesser extent to assess the feasibility of using a particular concept in a questionnaire, to determine if the questions seem to elicit appropriate responses, and to suggest additional questions or response categories which can be precoded on the questionnaire.

Other relevant objective information, which might affect the final questionnaire design, also can be obtained in an informal test--e.g., a

preliminary indication of the interview length (called respondent burden by OMB), and refusal problems.

Selecting the Sample and Site for the Test

Usually, adequate subjective information can be obtained from 50 to 300 respondents. The respondents generally are selected purposively rather than randomly to achieve the desired objectives of the test. For example, if the survey will be conducted with a general population sample, representatives from a broad range of subpopulations should be included in the informal test. On the other hand, if the questions being tested are directed at a specific subpopulation such as food stamp recipients or high income persons, the entire test sample might be composed of representatives of that group to ensure adequate coverage with a small number of interviews. When this is the case, the site selection may depend on the location of the subpopulation or availability of high quality records for use in selecting a sample. If no such constraints exist, then convenience and low cost are the chief factors in selecting a location, which frequently results in the selection of a site near the office conducting the work. Since informal tests are restricted to a few sites, the inferences drawn from the results do not necessarily apply to other sites or the Nation as a whole.

Considering the Design Issues

The study design for informal tests is probably more important than the number of interviews because subjective evaluations are not always improved by the quantity of observations. However, compared to formal tests or the actual survey, the design of an informal test is usually relatively simple. In planning for one, the following factors should be considered:

a. The questionnaire composition

A decision should be made on whether to test the entire questionnaire or only a portion of it. If only one test is planned, it is advisable to use the entire questionnaire since responses can be affected by the presence and order of the questions included in the proposed questionnaire. For this reason, questions borrowed from other surveys should not be omitted from this testing.

When a series of tests is planned, one or more of the informal tests may be devoted to a particular portion of the questionnaire that is expected to be troublesome. In such situations, the section tested might be relevant only for a particular subpopulation and the sample for the test might be limited to the population subgroup as discussed previously. At the end of the process, the entire questionnaire will have to be tested to see how the sections work together.

Another questionnaire choice concerns the possibility of using two or more versions of the question (or answer) wording or order. Although this is perhaps a more common technique in split-sample testing, it can be used effectively in an

informal test to make a quick comparison of the alternatives.

b. The interviewing method

Again, the choice of interviewing procedures is affected by whether a series of tests is planned. If the informal test will be the only test, the questionnaire probably should be administered in the same manner selected for the survey (e.g., self-administered, interviewer-administered in person or by telephone, or some combination of these methods). However, as part of a series in which the informal test will be used only for a preliminary indication, a different method may be justified to save time and/or costs.

c. The selection and training of interviewers (for interviewer-administered questionnaires)

There are advantages in selecting skilled, experienced interviewers for informal tests. With such interviewers, it is more likely that question misunderstandings or difficulties will be due to questionnaire design deficiencies rather than to the interviewer. They also can provide considerable assistance in improving the questionnaire based on their experiences with other surveys. However, there are some disadvantages also; e.g., they may be able to handle situations that will cause problems for less experienced interviewers in the actual survey. Thus, the use of interviewers with varying experience and skill levels may be desirable in an informal test.

The interviewers should know how to probe to obtain information that will be useful in refining the questionnaire. All interviewers do not possess these skills and should be trained on them, if necessary. They should understand the purpose of the test, and that they are expected to be critical of the questionnaire. Also, they should be thoroughly trained on the concepts and definitions used in the questionnaire, as well as on the proper way to administer the questionnaire. With a better understanding of the rationale and logic behind the questions, the interviewer should be able to make a more significant contribution to the evaluation.

Another option is for the questionnaire designers and researchers to serve as the interviewers. This ensures that the persons doing the interviewing are thoroughly familiar with the aims and objectives of the test. Questionnaire designers and researchers, who are inexperienced interviewers, should be trained on interviewing techniques before beginning their assignment; otherwise, they could adversely affect the test results. Even if they do not plan to perform this role, such training will make them more sensitive to the problems questionnaires can cause interviewers.

d. The observational feedback system

The most important element in the

design could be the system developed to capture subjective observations on the performance of the questionnaire in the informal test. There are several ways that this can be accomplished. For example, interviews can be tape-recorded, observers can accompany the interviewers and record information on a specially designed evaluation form, the interviewers can be provided with a similar evaluation form to be filled out, or the interviewers and/or observers can be debriefed following the test. Observers are extremely helpful because they can watch the interaction between the interviewer and respondent to detect problems that might not be apparent to the interviewer. In addition, interviewers and/or observers can initiate conversations with respondents at the conclusion of the interview to obtain a respondent's impressions about certain questions, and to identify questions that may have been troublesome to the respondent.

2. THE OPERATIONS AND EVALUATION OF THE TEST

Operational Issues

The evaluation of an informal test involving personal or telephone interviews can be hindered if steps are not taken to ensure that the questionnaire is administered properly. The persons conducting and observing interviews should understand the objectives of the test and the importance of not arbitrarily varying the questionnaire wording and administration. However, they should know how to probe by rewording questions or asking other questions when it is suspected that a response is inaccurate, inappropriate, or insufficient. Probing should only be used under circumstances approved by the questionnaire designer/researcher to provide further insight into potential questionnaire problems; when used, it should be noted as part of the feedback system.

The lines of communication between the questionnaire designers, observers, interviewers, and other project staff should be well established to enhance the feedback. One major advantage of an informal test is the possibility of making "on the spot" revisions to the questionnaire as a result of the feedback. Because of the small number of people and questionnaires involved, any problems uncovered can be discussed at the end of a day's interviewing and changes made before interviewing begins the next day. These changes and the rationale for making them should be carefully documented for use in evaluating the questionnaire's performance and for future use by others who are performing related work. Personnel involved in the evaluation should actively participate in the operational phase of the test.

Evaluating the Results

Much of the evaluation in an informal test is simply the use of common sense in reacting to problems identified by the feedback system. The lack of objective criteria for evaluating the questionnaire responses may be seen as a disadvantage of this type of testing. However, some quantification of the responses may be possible (e.g., tabulations of the number of Don't Know, Refused, or Not Applicable responses to a question). These types of responses in addition to inconsistent and missing responses

often identify various questionnaire problems. These tabulations can usually be performed clerically because of the small number of cases.

Unfortunately, the test often only indicates that there is a problem; it does not provide the "correct" solution. For example, if a given question is not answered frequently in a test, there may be a problem with the wording. However, unless the interviewers or observers have probed to find out why the question is not being answered, the questionnaire designer might not have enough information to rephrase the question in a way that will elicit more responses. Thorough documentation of the evaluation process and resulting questionnaire changes should be made for use by future researchers. Because the results are not statistically conclusive, informal tests are frequently the first step in a process leading to formal tests from which more reliable inferences can be made.

3. TIME, COST, AND STAFFING CONSIDERATIONS

Time

The amount of time required to conduct an informal test varies according to a number of factors. Assuming that the questionnaire has been drafted [1], the total amount of time that should be allowed for the operational aspects of an informal test is approximately three to four and one-half months. This includes time for OMB approval [2] (during which manuals, training, and field procedures can be prepared, an interviewing site and a sample of respondents can be selected, and forms can be designed if necessary), selection and training of interviewers, reproduction of necessary materials, data collection, receipt of feedback through interviewer debriefings, completion of observer reports, etc., and summarizing the results. The variable factors that prohibit specification of an exact time frame include: 1) the number of cases and interviewers; 2) the length of the interview and the distance between sample households; 3) whether materials can be duplicated in-house or must be sent to a printing company; 4) whether interviewer instructions, training materials, debriefing guides, and observer forms are written (the larger the number of sample cases, the more likely it is that these materials will be put in writing); and 5) whether materials have to be mailed to the interviewing site. The final schedule should always allow some time for possible delays due to unexpected problems; otherwise, the next draft of the questionnaire may be required for another test or for the survey before the results of this test are available.

Costs

Relative to other types of field tests, informal tests are inexpensive data collection efforts. This, in addition to the relative speed with which they are conducted, contributes to their usefulness as tools for questionnaire design.

It is difficult to quantify a cost range for conducting an informal test; a very tentative estimate would be \$5,000 to \$30,000 (1983 dollars) for a test for a large-scale national survey. However, the factors that contribute to the costs include 1) interviewing and field staff salaries (this is the major cost); 2) other

professional salaries (i.e., questionnaire designers, observers); 3) travel and expenses for interviewers and observers; 4) forms design and/or reproduction of questionnaires; and 5) postage (if materials need to be mailed to the field).

Staffing

Several types of skills are necessary to conduct an informal test, some of which may be combined in a single person. However, usually it is necessary to have a team of persons or several different groups of people.

If a team of persons is involved, someone must coordinate all the activities involved in the field test. This includes selection of the test site and sample, selection and training of the interviewers (if applicable), development of the questionnaire and a system for receiving and evaluating feedback about it. Experience with or knowledge of data collection operations is an essential qualification for this person. The qualifications to be considered in selecting interviewers have already been discussed. No special skills can be identified as requirements for the other staff members. However, the person(s) carrying out the evaluation of the test results should have the ability to recognize problems during an interview, or in a review of the completed questionnaires or tabulations, and ability to recognize the implications of the test results for the design of the questionnaire.

4. INFORMAL TEST EXAMPLE

Introduction

An informal test was conducted prior to the 1980 National Survey of Fishing, Hunting, and Wildlife Associated Recreation (FHWAR) to refine the proposed questionnaire. Although this survey had been conducted at 5-year intervals since 1955, it was acknowledged that the previous questionnaires contained some weaknesses. Specifically, there were needs for better data on "nonconsumptive users" of wildlife resources. Whereas many of the questions for fishers and hunters had been used in the previous surveys, the questions for wildlife photographers, bird-watchers, and other observers of nature were relatively untested. The Bureau of the Census conducted the test and the survey for the Fish and Wildlife Service (FWS) of the Department of Interior. This test was selected as an example here because it points out several different types of questionnaire problems that can be detected during informal testing (see the results section for a description of the problems). The cost of this test was approximately \$20,000, and it took a little over 3 months to plan, conduct, and evaluate.

Technical and Operational Considerations

The test was designed to use the basic methodology proposed for the survey, namely, a telephone screening interview with a household respondent, which was followed by a detailed personal interview with each household member who was identified as a hunter, fisher, or nonconsumptive user. Three questionnaires were used in this process: 1) a screening questionnaire to identify persons for further questioning; 2) a detailed questionnaire for hunters and/or fishers; and 3) a detailed questionnaire for nonconsumptive users. Persons who were both hunters/fishers and nonconsumptive users were administered both detailed questionnaires.

The methodology for the test varied from the survey in that a judgmental (nonprobability) sample was selected to provide a sufficient number of participants for personal interviews. The survey used a probability sample. The sample was selected from a list of respondents who had been in a survey conducted by the Michigan State Department of Natural Resources in 1979 and who were licensed to hunt or fish at that time. It was assumed that it would be impossible to reach many of these persons by telephone (wrong number, no answer, etc.) and that some of those reached would not be identified as hunters, fishers, or nonconsumptive users. Also, of those identified, some would be unavailable for a personal interview. Therefore, approximately 400 persons were initially selected from the list to ensure that at least 100 persons would be identified for a detailed interview.

Ten experienced Census Bureau interviewers were selected to enable the test to be completed within 5 days. A self-study guide was sent to interviewers to familiarize them with the concepts and procedures that would be used in the test. Then, classroom training was held to discuss the test procedures and provide practice in administering the questionnaires in mock interview situations. In addition, the Bureau prepared a Reference Manual to assist the interviewers in administering the questionnaire.

To aid in the test evaluation, Bureau and FWS staff members accompanied the interviewers to observe and report on the detailed interviews. In addition, the interviewers were encouraged to report any problems in a debriefing session following the interview period. The questionnaire data were not processed; however, some clerical tallies were made for evaluation purposes.

Results

The test results indicated that the screening interview could be used to identify hunters and/or fishers and nonconsumptive users who were eligible for the detailed interview. There were two major findings, based on subjective evaluations, regarding the adequacy of the screening questionnaire. First, it was observed that length was affecting cooperation. In the test, 10 out of 100 respondents refused to allow a personal visit interview because of the time it had taken to complete the screening questionnaire. Therefore, it was recommended that the screening questionnaire be shortened by dropping several questions that were unnecessary for screening purposes.

The second major finding was that although household respondents seemed to be able to identify hunters and fishers, they had more trouble identifying nonconsumptive users. It was thought that the loose definition of nonconsumptive users might be the cause; therefore, it was recommended that those screener questions be clarified.

The observers and interviewers thought there were several problems with the detailed questionnaires used in the personal interviews based on their subjective observations. In general, the questions seemed repetitious and wordy. To help the flow of the interview, changes in the interviewing techniques, skip patterns, and questionnaire format were suggested. Some

problems with specific questions included: 1) confusing wording; 2) deficient visual aids; 3) vague terms and concepts; and 4) missing answer categories. Again, appropriate improvements were suggested where possible. Clerical tallies of item nonresponses were also used to identify problems with specific questions and efforts were made to change the questions to elicit more answers. Also, it was believed that better interviewer training would have reduced the number of nonresponses in some of these cases.

Overall, it was noted that the structure of the detailed questionnaires led to potential double reporting of information; e.g., three reports of one trip that involved hunting, fishing, and nonconsumptive activities or three reports of the same trip by three family members who went as a group. On the other hand, trips originating from a vacation home were probably missed because of the wording of the introduction to this set of questions. This resulted in some suggestions for restructuring the questionnaire and rewording the introduction.

The revised questionnaire was used in the survey, which was completed in 1981. The FWS used the results to prepare a national report and individual state reports for the 50 States. The national report was released in November 1982, and the primary users, namely, fish and wildlife planners and managers at all levels of government, have found the data generally accurate and useful. These favorable results were probably due, in part, to questionnaire improvements arising from the informal test.

NOTES AND REFERENCES

- [1] The time required to draft the questionnaire varies considerably depending on how much developmental work is necessary--for example, whether the survey has been conducted before or is totally new, whether any developmental techniques such as focus group interviews have been used, etc.
- [2] OMB approval is required for all data collection efforts paid for by the Federal Government that will involve more than nine respondents. OMB's role is to ensure that information collected is in the public interest, that respondent reporting burden is reasonable, and that certain statistical standards are met. OMB now (1983) requires 60 days to review requests for approval.
- [3] Babbie, Earl R. Survey Research Methods. Belmont, Ca.: Wadsworth Publishing Company, 1973, pp. 205-221.
- [4] Courtenay, Gillian. "Questionnaire Construction." In Hoinville, Gerald, Roger Jowell and Associates, Survey Research Practice. London: Heinemann Educational Books, 1978, pp. 51-53.
- [5] Goode, William J. and Paul K. Hatt. Methods in Social Research. New York: McGraw-Hill Book Co., 1952.
- [6] Moser, Sir Claus, and G. Kalton. "The planning of social surveys." In Moser, Sir Claus and G. Kalton, Survey Methods in Social Investigation. New York: Basic Books, 1972, pp. 47-51.
- [7] Payne, Stanley L. The Art of Asking Questions. Princeton, N.J., Princeton University Press, 1965, pp. 13-14.