

WORKSHOP ON QUESTIONNAIRE DESIGN

Resource Paper

Question Formulation and Instrument Design

Arnim H. Meyburg

*School of Civil and Environmental Engineering
Cornell University*

Helen M. A. Metcalf

PlanTrans

INTRODUCTION, RATIONALE, AND BACKGROUND

Common sense should dictate in large part the design of survey instruments, once a decision about the appropriate survey procedure has been made. And it should also dominate the issue of question formulation. Nowhere in the whole travel survey methodology is it more true that the development of travel surveys resembles an art rather than a science. In the context of survey instrument design and question formulation, we are concerned with controlling possible bias that can be introduced through poor design. The discussion is outside the more tangible and precise statistical assessment of (sampling) error. However, the two major concerns, namely, that of undesirable variability [“instrument uncertainty”, (1), p. 148] in the data collected, and that of ending up with data that are not what were the intended data (“instrument bias”), are prevalent in survey research and practice.

It is a truism that anybody can develop a questionnaire. It is, however, a completely different issue whether such a questionnaire will generate the information that the developer and analyst are interested in obtaining, and that is suitable for answering the survey objectives. And this question needs to be asked with respect to the quality, as well as with respect to the quantity (nonresponse issue), of the data.

This resource paper focuses on the issue of question formulation and survey instrument design. No standardized procedures for question formulation and survey instrument design have been developed, nor is it likely that they will be developed in the near future, given the variety of survey objectives, client expectations, social, cultural, and language environments, etc. At best, guidelines that reflect lots of empirical evidence can be put together to provide practical guidance to the travel survey designer. One example of an extensive set of guidelines is presented in Richardson, Ampt, and Meyburg (1). Another example is the 1996 Travel Survey Manual, prepared by Cambridge Systematics, Inc., for USDOT (13). Other treatments can be found in more specialized contexts. The most recent U.S. example is being developed by the National Opinion Research Corporation (at the University of Chicago), in the context of nonresponse issues (14).

This paper has three main objectives: First, it examines some theoretical and methodological elements of the social science literature on data quality, as it relates to questionnaire formulation and survey design. Second, the paper provides discussions of practical aspects of question formulation and survey design. And third, it develops a set of succinct state-of-the-practice recommendations for improved question formulation and survey design.

DATA QUALITY IN RELATION TO SURVEY QUESTION AND DESIGN CHARACTERISTICS

A large amount of research concerning the quality of survey data has been conducted in the social sciences, especially in the fields of social and cognitive psychology. This section examines the aspect of response quality in relation to question formulation and survey design. While much of this work concentrates on determining validity and reliability of attitudinal survey data, the insights gained and the research methodology are applicable across all types of survey topics.

Scherpenzeel and Saris (2) have tackled in a systematic manner the validity and reliability of survey responses as a function of variations in survey question characteristics. While there is a plethora of anecdotal and empirical evidence that suggests certain fundamental guidelines about question characteristics, as documented later in this paper, Scherpenzeel and Saris (2) report on their research efforts in the Netherlands to overcome the lack of “general rules connecting the degree of measurement error with different combinations of these characteristics” (p. 341).

They acknowledge that for each survey, choices have to be made about question wording, response scale (especially for attitudinal questions), question context, and data collection technique. While each of these choices can introduce different errors, little is known about the magnitude of these errors and their effect on the validity and reliability of survey responses. Their 1997 study reports on how data quality is affected by a set of survey question characteristics. Their objective was to state in quantitative terms the relationship of measurement errors to these characteristics, in order to facilitate the prediction and optimization of data quality. Their study also “shows how relationships between variables can be corrected for measurement error on the basis of this information.”

They use so-called multitrait-multimethod (MTMM) models to first estimate the validity and reliability based on a substantial number of survey measures. Subsequently, in the “second stage” (or “meta-analysis”) of this approach, “the variation in these quality estimates is related to variation in the characteristics of the survey questions.”

Other researchers working on survey research in the fields of social and cognitive psychology have made earlier similar attempts to gain systematic insights into data quality and various aspects of question wording, context and sequencing, retrieval method, length of survey, etc. Specifically, Andrews (3) was the first to use the MTMM approach. Rogers, Andrews, and Herzog (4) built on this approach using North American samples, as did Költringer (5) in Austria. Scherpenzeel and Saris (6) also conducted an international comparison of data quality for an individual topic.

Another noteworthy effort along this line of enquiry is a 1991 study by Alwin and Krosnick (7) on the reliability of survey attitude measurement. They define reliability as the “normed linear relationship between two attempts to measure the same thing, net of true change.” This definition is based on Lord and Novick (8).

Alwin and Krosnick (7) tested 96 attitude measures from five U.S. national panel surveys. Their paper includes a rather important discussion of definitions of reliability and validity, terms that seem to be used with various interpretations throughout the literature. Given that the overall objective of systematic studies of attitude measurement error is to determine the degree of reliability and variability in the response data in order to improve, if not optimize, survey and survey question design, it is crucial that a high degree of commonality on these definitions be developed. Only then can practitioners actually draw

proper conclusions from the results of these methodological studies. And only then can we assume that optimization of design and determination of reliability and validity can be performed consistently across surveys. Otherwise, the methodological efforts to systematically determine reliability and variability across sets of surveys and question characteristics will not be much different from the frequently used basis of “anecdotal” empirical improvements from survey to survey.

The special significance of the 1997 study by Scherpenzeel and Saris (2) lies, in part, in the fact that their meta-analysis also includes the effects that were investigated by Andrews (3), Rodgers, Andrews, and Herzog (4), Költringer (5), and Alwin and Krosnick (7). In this respect, their research represents an advance over and a synthesis of the meta-analyses performed in a variety of contexts and cultural and language environments. They state that their research can be used “to construct optimal questions” and “to correct unavoidable measurement errors in survey research.”

One should caution, though, against transferring results of such analyses blindly across different language and cultural, as well as social and educational, environments. Scherpenzeel and Saris (2) themselves acknowledge these limitations and admit that the transferability (“generalizability”) is limited. They specifically state that they do not see its applicability beyond the Dutch language environment.

However, it should be clearly acknowledged here that their methodological research into reliability and validity of data resulting from attitude measurements is of substantial significance for others to build on in other language and cultural contexts. The comparison and the potential merging of research results, where commonality of approach and results warrants such actions, has far-reaching consequences for improving the state of the art of survey research and practice.

Tourangeau and Rasinski (9) address the question of “context effects” on attitude measurements. Based on an emerging agreement among social psychologists, they define attitudes as “structures in long-term memory”. Their analysis of the cognitive process that is required to answer attitude questions is based on the assertion that a four-stage process is involved on the part of the respondent. The four stages are assumed to be the following:

1. Interpretation of the question (“What is the attitude about?”).
2. Retrieval of relevant beliefs and feelings.
3. Application of these beliefs and feelings to generate appropriate judgment.
4. Use of this judgment to formulate response.

An understanding of this process is deemed necessary for the understanding of the reliability of respondents’ answers to attitudinal questions. Each of these stages is likely to be influenced by psychological variables that, in turn, may covary with the data collection method [Schwarz et al. (10)]. If the response is to be entered into a given set of response options, as is the case in many survey instruments, the translation of the response into one that appears on the list can distort the true attitude. Also, given the fact that the survey analyst may edit the response in line with consistency checks among answers, the final answer about an attitude may be quite different from the true attitude the respondent carries in his/her long-term-memory structure.

Furthermore, they claim that all four stages can be affected by “prior items”, i.e., certain conditioning of the respondents’ minds that occurred because of having answered related questions earlier in the survey. They are either building on that thought process in a

constructive way, or they may aim for “matching” the earlier responses for maximum consistency. The issue of social and political correctness, as well as the pressure to appear intelligent, consistent, rational, timely, etc., plays into this process in a significant way.

Answers to survey questions, particularly those that reflect attitudes and opinions, are subject to a number of influences, as discussed above and as well-documented in the literature. However, the cognitive processes that are at the root of these influences have received increased attention and have been subjected to systematic studies only fairly recently [e.g., Tourangeau and Rasinski (9)].

Tourangeau and Rasinski (9) showed that the retrieval method, i.e., the survey method, generated different pressures on the respondents with respect to time and social and political correctness. The retrieval process may produce the most easily recalled (“accessible”) aspect, rather than the most relevant and important for the question at hand. More careful, i.e., more time consuming reflection may be necessary to access the less-than-spontaneous and/or intuitive response. This reflection may not be possible under phone or personal interview pressure (p. 301). Hence, it is likely that a shortened or superficial retrieval process from the long-term memory bank takes place and generates an attitudinal response that is not reflective of the true attitude carried by the respondent. A survey suffering from these handicaps will result in a “sample of pertinent beliefs that overrepresents the most accessible ones” (p. 301).

Another aspect of potential unreliability of attitude measurements results from the fact that respondents simply may not have a well-formulated attitude in their belief structure. Hence, their response to attitudinal questions may easily be swayed by the way the issue is posed, the sequencing of questions, the particular wording of a question, and the item context. Schumann and Presser (11) have shown that such variations can have pronounced effects on the responses.

This aspect obviously has important implications with respect to the survey method used (issue of time, social and political correctness pressures) and the sequencing of attitudinal questions (“prior items” issue). Ignoring these aspects will certainly affect the reliability of attitude measurements.

Schwarz et al. (10) demonstrated the importance of a “psychological approach to survey methodology that incorporates cognitive and communicational variables.” Specifically, by means of experimental evidence, they investigated question answering in relation to different survey methods (self-administered, phone and personal interview). They found significant differences in response data quality, depending on the survey method, with respect to question wording, recall period, and question order. Also, they demonstrated the well-known interviewer effects on response quality. For a discussion of a structural modeling approach (including MTMM models) for the evaluation of measurement instruments, see, for example, Saris and Andrews (12).

FACTORS THAT AFFECT INSTRUMENT DESIGN

Question formulation is very much dependent on the objectives of the survey and the target population to be surveyed. These preliminary, yet fundamental, survey planning considerations give rise to decisions about the most appropriate survey method to be employed, which, in turn, determines the most suitable instrument design (1).

Survey Objectives

The basic objectives of a survey may be self-evident; for instance, the main objective of most household travel surveys is to collect information about how and why people travel or do not travel during a certain period of time. But a more specific, and more important, objective may be to collect particular information for use in modeling procedures. Secondary objectives may include the desire to create a survey that can be easily compared with other household surveys. The objective of many transit surveys is to collect information about how people use the local transit system, or to collect attitudinal information to be sure that the service is meeting the needs of the customer. In most cases, no matter what type of survey is to be conducted, the objective of the survey will define the target population to be surveyed.

Target Population

The definition of the target population, or sample, may be as simple as “anyone who travels,” or as complex as “people who began riding peak buses on a daily basis in a particular region within the last six months.” In most surveys, a secondary definition will likely include a restriction on the age of potential respondents. But whether the definition of the sample is simple or complex, the population selected will partially or completely determine the method of collecting data. This is because the method of collecting data from a population is dependent on the ease of locating that population. For instance, when conducting a survey where the objective is to collect origin-destination data from transit riders, the easiest method of finding transit riders in most regions is by intercepting them at transit stops or stations. It would not be wise to conduct a random-digit household telephone survey of transit riders if the incidence of transit riders in the region to be surveyed is 3%. Once the target population has been identified, it is crucial to the success of the survey to identify the easiest and most efficient way of locating and surveying them.

Survey Methods

There are a number of survey methods that can be employed for most surveys, among them face-to-face interviews, mail-out/mail-back surveys, onboard self-administered surveys, telephone interviews using paper surveys, computer-aided telephone interviews (CATI), computer-aided personal interviews (CAPI), a combination of these, and others. Basically these methods involve paper or computerized survey instruments that are either completed by an interviewer in the presence of the respondent, completed by an interviewer who is in contact with the respondent over the telephone, or completed by the respondent on his/her own. As explained in the previous section, choosing the appropriate method has much to do with the objectives and the target population of the survey. However, a number of other factors also contribute to this decision, such as the resources available, the time frame in which the survey must be completed, the literacy of the population to be surveyed, and the desired number of completed surveys. For instance, a CATI survey would not be a wise choice for most small-sample surveys, since the development and programming of the questionnaire take a significant amount of time and money. Although there are no standard guidelines for the size of CATI surveys, it would be safe to say that a survey of fewer than 2,500 would not be a good candidate for CATI,

while a survey of 5,000 or more would definitely benefit from CATI. Researchers who have experience with the costs of both paper and CATI surveys (and these costs will vary from region to region and from company to company) will be able to pinpoint the number of surveys at which one method is more cost-effective than the other.

Basic Instrument Design

There are numerous resources available to the survey designer, including an extensive list of guidelines for proper physical design and layout for personal interview surveys, self-completion surveys, and intercept surveys, provided by Richardson, Ampt and Meyburg (1). The most fundamental consideration when beginning the design of a survey instrument is whether or not the respondent is going to see the survey instrument. The following sections discuss different aspects of survey design, depending on whether the survey is to be self-administered or interviewer-administered. Additionally, there are some unique aspects of CATI surveys that should be utilized by the questionnaire designer.

Self-Administered Surveys

Self-administered surveys require special attention to appearance, since they will be placed in the hands of potential respondents. A self-administered survey is usually mailed to the respondents, or handed to them, for instance, on a bus. In the case of self-completion surveys, the professional appearance, or lack thereof, can literally determine the fate of the questionnaire during the first minute of encounter with the potential respondent. Of particular concern are:

- The length of the questionnaire,
- The spacing of questions,
- The size of typefaces used,
- The layout of questions and answers,
- A logical order to questions,
- Minimized writing,
- Simple language, and
- The provision of survey/question instructions.

The *length of a questionnaire* is often a factor of the resources available for a survey, and is more broadly defined by the objectives of the survey and the analyst's desire to know, rather than by the potential respondent's willingness to provide the type and amount of information requested. Experience shows that the length of a questionnaire has a significant impact, both on the quality of the information gathered, and on the willingness of the respondent to complete the survey. It is obvious that if a survey booklet is placed in a respondent's hands, the respondent will likely react differently than if he/she were given a survey the size of a 3" x 5" card. It is imperative, therefore, that only questions that are absolutely necessary for answering the objectives of the survey be included. This is especially important for self-administered surveys, since the respondent can see at a glance what is being asked of him/her. A secondary issue relating to surveys that have to be mailed out or mailed back is postage, which could seriously impact a survey's budget if the questionnaire consists of several pages.

The *spacing of questions* is also a concern for self-administered questionnaires. Related to the perceived length of the questionnaire is the question of whether to create a small, compact survey or a longer, more openly spaced survey. Generous spacing results in greater ease of writing responses, and allows *the use of larger typefaces* for easier reading, while at the same time adding to the bulk of the survey instrument. The size of the typeface used is of particular importance for collecting responses from the elderly, and if the population to be surveyed contains a large number of elderly persons, designing a survey instrument in a large typeface is essential. Most often, the answer lies in a compromise between compactness and ease of use. Only pretests will provide the definitive answer as to what approach is preferable for the specific target population to be surveyed.

Another item of consideration for self-administered surveys is the *layout of questions and answers*. A good layout not only looks pleasing, it also can help to utilize space more efficiently and should make it easier for the respondent to easily identify which answers correspond with which questions. In Figure 1 below, Example A illustrates a good layout design of the question, while Example B shows a poor one. In Example A, the question is set apart from the answers, the answers are lined up next to the tick-off boxes, and the box around the entire question will help to distinguish it from questions 5 and 7.

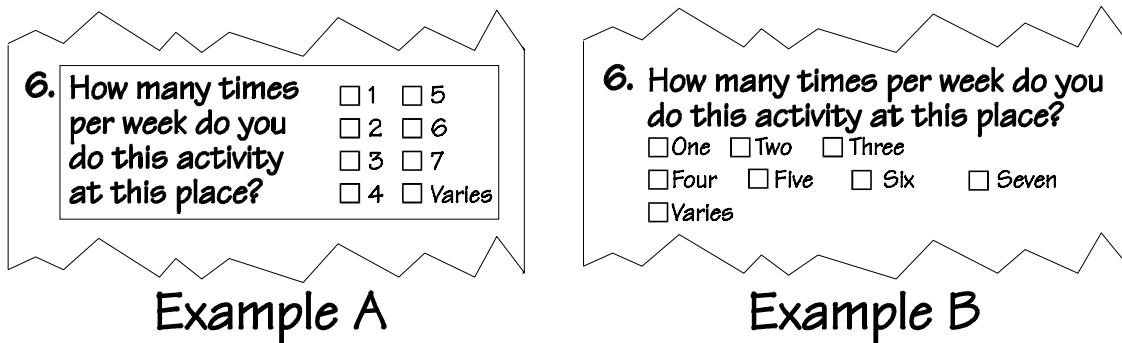


FIGURE 1 Examples of question-and-answer layout.

In Example B, the question and answers run into each other; the answers are more difficult to quickly identify (as opposed to the answers in Example A); the answers and boxes are not lined up, making them look messy and taking up more space; and the question will not be easily distinguished from the questions that come before and after it.

The guidelines for *logical question ordering* are fairly simple and obvious. The guiding principle ought to be to arrange the questions in categories, where both the ordering of categories and the sequencing within categories follow a logical and intuitively obvious path. It should avoid forcing the respondent to make mental jumps, which would give the appearance of a haphazardly composed survey instrument. Using boxes to surround each block of questions, headings, and even a sentence of introduction to each section can be helpful.

It is extremely important that a *minimum amount of writing* be required of the respondent to a self-administered survey. Questionnaires containing many open-ended

questions will cause a writing burden on the respondent and make processing completed questionnaires difficult and time-consuming. Even in closed questions, an obvious way of minimizing the amount of writing required is to provide tick-off boxes, lined up in columns next to the most frequently expected responses. Using *simple, non-technical language*, avoiding jargon and slang, has to be an overriding principle of question wording, as is the use of short and simple words. Obviously, many respondents will not know anything about a subject that is technical, and will not be familiar with corresponding technical words, phrases or acronyms. And while it is important to avoid using language that appears to be insulting to some respondents' intelligence, it is equally important to remember that the sample of respondents will presumably represent a cross-section of the population, including all levels of literacy. Using short and simple words and sentences also makes a questionnaire easier to read, and therefore may cut down on the amount of time it takes the respondent to complete the survey. For instance, using "How did you get here?" is easier on the respondent than "How did you arrive at your destination?"

Another item of consideration for a self-administered survey instrument is the inclusion of *survey and/or question instructions*. Whenever possible, it is advisable to place instructions directly in the survey instrument, as opposed to providing a separate sheet of instructions, which could be totally overlooked. In a booklet survey, an entire page in the front of the booklet could be devoted to instructions, but too many instructions can also overwhelm respondents and encourage nonresponse. Items of particular importance, such as when to complete a travel diary and what to do with it when it has been completed, can be provided in a prominent place in a large typeface. One good strategy is the provision of a filled-in replica (possibly miniaturized) of the survey forms, particularly for time-sequenced travel and/or activity reporting sheets. Occasionally, it may be necessary to provide instructions at the beginning of a block of questions, or perhaps preceding a particular question. However, since many respondents will *never* read instructions, good survey design practice suggests that the questions ought to be formulated in such a manner that they are unambiguous and easily understood, even by the more moderately literate respondents.

Interviewer-Administered Surveys

Surveys that are administered by an interviewer are usually completed over the telephone or in person with the respondent. While the designer of a self-administered survey needs to be concerned with what the respondent sees in his/her survey, the designer of an interviewer-administered survey needs to pay particular attention to what the respondent hears. Of particular concern to the designer is:

- The length of the questionnaire,
- The introductory paragraph(s),
- Minimized writing,
- Simple language, and
- The provision of an interviewer manual.

Although the *length of a questionnaire* may not be initially obvious to a respondent over the telephone, it is still important to avoid developing an overly lengthy questionnaire.

While a lengthy questionnaire may cause respondent fatigue and result in the respondent hanging up the telephone before the survey is completed, overly lengthy surveys that will be conducted long-distance can seriously impact the cost of the telephone calls. Surveys that are conducted face-to-face should also not be overly long, since the interviewer may have trouble physically handling the long survey, and because the respondent may be able to recognize that the survey will be lengthy and may decline to participate. Length is also of particular concern if the survey involves intercepting potential respondents at transit stops or stations, since the interviewer may have a limited amount of time between intercepting the respondent and the arrival of the bus or train.

Thoughtful wording and sequencing of an *introductory statement* in an interviewer-administered survey will significantly affect the interviewer success rate within the very first few moments of a phone call or intercept. Thus, the opening paragraph of an interviewer-administered survey needs to be designed in such a manner that trust is built quickly, before losing the contact. It is a requirement for success to put the respondent at ease about the legitimacy of the survey, the importance of the subject matter to the respondent, and the sincerity of the interviewer. Only after these essential steps have been accomplished is it feasible to expect success in the actual survey questions. Hence, it is important to stress to interviewers the importance of conscientiously adhering to the wording of an introductory paragraph. It may also be helpful to test several opening statements during a pretest, to determine which one provides a better response.

Interviewer-administered surveys should require a *minimum amount of writing* or computer input on the part of the interviewer. Similar to self-administered surveys, questionnaires containing many open-ended questions will cause a writing burden on the interviewer, cause the respondent to spend time waiting for the interviewer to write or type, and make processing completed questionnaires difficult and time-consuming. As in self-administered surveys, use of *simple, nontechnical language*, avoiding jargon and slang, and use of short and simple words and sentences are important in interviewer-administered surveys, however, for different reasons. An interviewer who trips over long sentences, multisyllabic words or technical jargon is in danger of appearing unprofessional, which could compromise the trust of the respondent. Additionally, if a respondent cannot understand what the interviewer is saying, he/she may abandon the survey before it is complete, or he/she may pretend to understand and guess at the answers because of fear of embarrassment.

Although it is not part of the survey instrument, one item of particular importance for the interviewer-administered survey is an extensive *interviewer instruction manual*. This is essential, both as part of the interviewer training program and for reference purposes, should the interviewer become unsure about proper conduct and proper answers to respondents' questions in the course of the survey. The interviewer manual should review the basic purpose of the survey; indicate the sponsor or agency for whom the survey is being conducted; provide a name and telephone number, as appropriate, for respondents to call to check on the validity of the survey or ask questions; and discuss any issues of particular importance. The survey developer should review the completed survey instrument and determine what questions might cause particular problems for respondents or interviewers. A survey pretest is particularly helpful in determining what kind of assistance the interviewer may need, or what survey questions are particularly bothersome to respondents.

Computer-Aided Telephone Interviews (CATI)

While CATI surveys are interviewer-administered and can benefit from all of the items mentioned in the previous section, there are some special items to consider for CATI surveys. Although CATI can be used for many types of surveys, this section will refer to CATI in terms of a household travel survey, which is generally a more complex survey that can very effectively utilize the benefits of CATI.

The computerized aspect of CATI provides a number of tools to the questionnaire developer that should be utilized very carefully. Specifically, the following items are very important:

- Developing of complex skip patterns,
- Double-checking responses,
- Referring to information already collected, and
- Testing the programming.

One tremendous advantage of CATI surveys is the ability to build *complex skip patterns* into the survey. For instance, respondents are to be asked two questions about travel from “Point A” to “Point B”. Question 1 asks the respondents what mode of transportation they used. Question 2 is actually three different questions, depending on the respondents’ answer to Question 1: If they traveled by car, the interviewer wants to know where they parked it and how much it cost; if they traveled by bus or by train, the interviewer wants to know how they paid their fare and how much they paid; and if they traveled by bicycle, it is important to find out what they did with their bicycles when they got there. Using CATI means that the computer can skip directly to the appropriate Question 2 based on the answer the respondent gives to Question 1, without the interviewer having to identify which question to ask.

CATI is also capable of significantly reducing erroneous data, because a well-written CATI program will *double-check responses*. For instance, we ask for the ages of each member of the household, and we ask how each member traveled during the travel day. If a household member traveled by car, we ask whether he/she was a driver or a passenger. If a household member indicates that he/she was the driver of a car on a particular trip, the computer can double-check the age of that member, to be sure that the answer is plausible. Without CATI, it would be the responsibility of the interviewer to remember that 30 questions ago he/she was told that this household member was 5 years old. With CATI, the computer program can identify the problem immediately and have the interviewer verify the age of the household member and the method of travel, and make necessary changes. Probably the most difficult aspect of using this benefit of CATI is locating all the answers in a survey that can be double-checked.

A CATI survey can also refer to *information that has previously been collected*. In a household travel survey, this aspect of CATI is particularly helpful. For example, suppose a household travel survey collects detailed information about the vehicles that are available to a household (e.g., make, model and year), and then asks the respondent to indicate which vehicle was used for each trip taken on the travel day. Once the data on each vehicle are collected, a CATI program can present a list of the vehicles for the interviewer to select from for each trip taken in the household. Figure 2 provides an example of this.

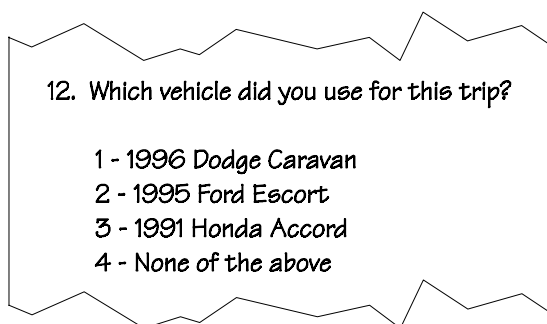


FIGURE 2 Example of a CATI vehicle list.

Since many people use the same vehicle throughout the day, once a household member has indicated the use of a particular vehicle for the first trip of the day, the CATI program can place that information into subsequent questions, for example, as shown in Figure 3.

Adding information that was previously collected into the text of subsequent questions and answers is a very useful way of making a survey easier on the respondent and the interviewer.

Finally, it is absolutely necessary that the CATI survey *programming be thoroughly tested* to ensure that the program works correctly. Adequate testing involves checking the program's responses to every possible answer to each question, and making adjustments as necessary. Since it may be impossible for one person to think of every possible answer, having several people check the program is often worthwhile; however, this type of testing is not a substitute for a pretest. It is also very important to test the program's ability to "back up" several questions. Most CATI programs allow an interviewer to go back to a previous question by backspacing through the survey one question at a time. This can be necessary when respondents change their mind about an answer given previously in the survey. The programming that controls skip patterns, references to previously collected data, and double-checking can interfere with a program's ability to backspace, change an answer, and then move forward again accurately. Therefore, careful checking is essential.

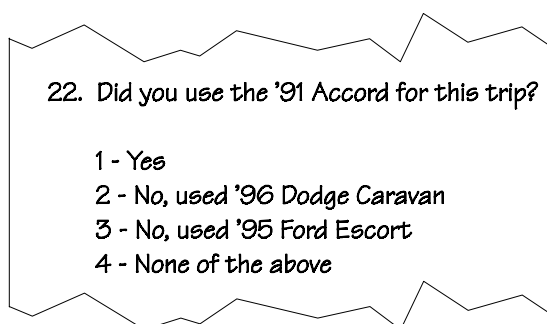


FIGURE 3 Subsequent CATI question.

Question Types

Surveys typically contain one or more of the following question types:

- Classification questions,
- Factual questions,
- Opinion and attitude questions, and
- Stated response (preference) questions.

Classification questions have two main purposes: first, to gather data that will allow subsequent stratification of the sample for analysis of specific population groups; and second, to be used as a screening device for determining whether a respondent is of interest to the survey objective. Thus, rare populations can be identified, even when there is no prior sampling frame available for them.

Classification questions can also be used as branching questions to ensure that respondents are only asked questions that are relevant to them. Classification questions need to be formulated with particular care, since the reason for their existence may not be obvious from the stated survey objectives, and they may appear to be intrusive to the respondent. This latter point is particularly significant, since these questions typically appear at the beginning of the survey, and respondents may not have developed that kind of commitment and involvement in the survey subject matter, so that it is still easier for them to terminate the interview or the completion of the questionnaire.

Factual questions are those that record actual events, and experiences, i.e., “facts,” that the respondent is expected to know. The survey designer must make absolutely sure that the questions do not include technical jargon unfamiliar to the average person, so that any non-reporting would be a consequence of a misunderstood question, rather than of a fact, e.g., an activity, not having taken place. Furthermore, since these questions solicit reporting of actual events, not of average expected events or facts, it is important to make sure that the time horizon over which these events are to be recalled is well within the typical memory capability of the average respondent. The inability to recall, if the time horizon prescribed by the survey is too long, will lead to potentially severe distortions of actual behavior, e.g., travel, and, hence, to potentially disastrous consequences when the survey results are used in support of transportation investment policies that, by nature, are rather costly. Several studies have investigated this effect, e.g. Clarke, Dix and Jones (19), Meyburg and Brög (20), and Stopher and Sheskin (21).

Opinion and attitude questions constitute the most challenging of question types in a survey. The reason is obvious: since opinions and attitudes, rather than facts, are solicited, even the slightest suggestion in the way a question is formulated can potentially lead the respondent toward a particular answer. This is particularly true in personal or telephone interview surveys, where the script needs to be followed verbatim and with consistent intonation.

In questions soliciting attitudes and preferences, care must be taken to account for some respondents’ propensity to answer in a manner that is intended to please the interviewer or that makes the respondent appear in the best possible light, regardless of what his/her true preferences and attitudes are. Interrupting repetitive question patterns and allowing an explicit unstructured response may serve as checks of the reliability of the information provided.

The complexity of the topic is reflected, in part, by a large body of literature dealing with opinion and attitude surveys, and question formulation in those types of surveys.

Stated response (preference) questions are a special case of attitude and opinion questions. They represent a multidimensional scaling technique, most typically used in travel choice situations. The analyst presents the respondent with a series of hypothetical combinations of attribute levels, and asks the respondent to rank them in order of preference, assign a rating to each combination, or choose between paired comparisons for a series of two-way choices. Of course, another option is to simply ask for the top choice from among the alternatives presented to the respondent.

When hypothetical questions are used, it is imperative to include cross-referencing to these questions, in order to determine how committed respondents are to their answers, and also to show consistency in preferences and attitudes. Another technique for making the answers to hypothetical questions useful is to include meaningful trade-off questions that test respondents' preferences in view of a sacrifice or concession to be made elsewhere.

One typical pitfall of this category of questions, aside from the conceptual drawback inherent in the fact that they pose hypothetical situations, is that, frequently, too many combinations and choices are presented to the respondent, who simply finds it impossible to make meaningful assessments that will subsequently allow the analyst to draw conclusions about true preferences for attributes or attribute combinations pertaining to travel choices. Also, while much work has been done toward developing sophisticated experimental designs, a perennial problem remains the quality of the questions, so that they remain understandable to the respondent. Bradley and Daly (22), among others, have addressed this issue in a systematic manner.

Answer Formats

There are three main ways to format the answers to questions: open-ended, field-coded, and closed. Many questionnaires effectively utilize all of these formats, but it is important to understand the advantages and disadvantages of each format.

Open-ended questions are those that allow respondents to write in or speak their answer freely, without having to choose a predetermined response category. Open-ended questions are useful for allowing the respondent to express opinions, attitudes, or preferences, and they are appropriate in cases where the response categories do not fit into a concise pattern or set that can be anticipated by the survey designer. Hence, they can be used effectively in pretests designed to develop proper response categories for the main survey. While open questions are likely to reveal interesting facts about and opinions of the respondent, they can constitute a real challenge for the analyst with respect to their coding and evaluation. Furthermore, the quality, detail, and length of answers are very highly correlated with the respondent's literacy level and propensity to give rather long-winded answers. Hence, open questions should be used with discretion and caution. On the other hand, if an open question has the effect of making the respondent feel that his/her opinion is of importance, the existence of such a question can have a constructive effect on the quality and completeness of the remainder of the response.

Even in surveys where open questions are not deemed necessary or appropriate, it is important to have a "Comments" space for respondents (both in personal and phone interviews and in self-administered questionnaires), for some of the same reasons that were identified in the preceding paragraph.

Field-coded questions pertain to interviewer-administered surveys. In this case, the question appears to be an open-ended question to the respondent, while the interviewer extracts the essential information from the “open answer” to record it into predetermined categories. While this procedure has the obvious advantage that it reduces the need for post facto processing, it clearly is subject to potential bias introduced by the interviewer, who determines the appropriate answer category. The advantages, of course, are the same as those identified for open questions in self-administered instruments.

Closed questions, which provide predetermined response categories, are and should be the most frequently used type of question, because they leave the decision up to the respondent as to where an answer fits. They record factual information, as the respondent sees it. At least two important guidelines need to be observed here. First, sufficient response categories need to be provided to ensure that the respondent is likely to find the appropriate answer. With the exception of “yes/no” questions, it may be nearly impossible to identify all possible answers to closed questions. Therefore, including “Other, please specify” in most closed questions is a good idea. This is particularly important so as to avoid forcing a respondent to choose an answer that is not at all appropriate. Second, great care must be taken to formulate closed questions in as unambiguous a manner as possible, in order to avoid respondent frustration and bias.

In order to avoid confusion, it may be necessary to provide very specific answer options. For example, in many self-administered household surveys, respondents will incorrectly indicate that they went to work or school, when they actually dropped off somebody at work or school. As shown in Figure 4, this mistake can be avoided by providing a set of more specific responses: pick up or drop off a person at work, pick up or drop off a person at school, and pick up or drop off a person at other. A source of ambiguity in answer categories may arise from a lack of specificity. In some cases, providing a definition in the answer can prevent the problem, such as defining a full-time worker as working 30 or more hours per week.

1. What was the next thing you did?
(check ONE only)

At home activities

Pick Up or Drop Off a Person { At work
At school/childcare
Other

Work
Work-Related

School/College/University
Childcare (day care/after school care) ...

Shopping
Social/Civic/Church/Recreation

Banking/Personal Business
Eating Out
Medical
Other (please specify) _____

FIGURE 4 Examples of specific responses.

The order in which the answer options are provided can be critical in obtaining the correct responses, as well. In general, the broader and more-common answer categories should appear near the end of the list, and the more-specific and less-common ones near the top, since the respondent is likely to stop checking a list if an apparently applicable answer has been encountered. An example of this is provided in Figure 4, where “at work” appears after “pick up or drop off a person”; while the “pick up or drop off” answer is a less common activity, listing it first may prevent many respondents from checking the box for “at work”, before they get to the “pick up or drop off” option. Similarly, since respondents frequently will not think of them as modes of transportation, it is prudent to list less-frequent modes like walk and bicycle first, so that respondents do not ignore or miss them.

Other Items for Consideration

This section discusses a number of items that can be problematic or should be seriously considered during the survey development process. Among these are:

- relevance of the questions,
- language problems,
- vague words,
- loaded and leading questions, and
- sensitive questions.

It is important to consider the *relevance of the questions* from the respondents’ point of view, as well as in relation to the survey purpose. All too often, the analyst, being deeply involved in the survey purpose and survey design, loses track of the fact that the respondent typically perceives the survey in a totally different frame of mind. A related issue is that of the reasonableness of a question, i.e., can we reasonably expect the respondent to answer either in fact or in terms of an opinion, attitude, or preference?

One effective technique for testing the adequacy of the instrument design with respect to content, length, and relevance of the questions was proposed by Oppenheim (17), namely by running through the survey stages in the reverse order. This means conducting tabulations and analyses with the expected, albeit hypothetical, data derived from the survey. This test tells, among other things, which questions are essential, and which are unnecessary.

If the survey sample consists of a specific sociodemographic segment of the population, the wording should be adjusted to take the population characteristics into account. This guideline pertains especially to situations where the respondents’ native language is most likely not the national language. If travel characteristics and travel behavior of specific nonnative speakers are the subject of the investigation, then the use of the appropriate foreign language is obviously mandatory (e.g., in the case of increased transit use in New York City, attributable to a recent immigration wave from the former Soviet Union.) Of course, the principles of question wording discussed here are pertinent to any language environment.

Stopher et al. (18) developed a creative onboard survey design solution in order to overcome previously experienced very low response rates due to rider disinterest and severe language barriers in Dade County (Florida) bus ridership surveys. It is an interesting example of successful data collection under difficult conditions, where the survey design was instrumental for the success in generating a data set of sufficient size and quality to

make it useful to the transit operator for evaluating service improvements. The language problem was overcome by virtually eliminating the need for verbal or written communications through the use of color-coded cards, whose distribution and collection were carefully monitored to provide the correct trip counts along the selected bus routes. Pictorials, prominently displayed on the bus, “explained” the card distribution and collection procedure to the riders. The problem of cooperation was resolved by making the whole process simple, and by the survey staff’s maintaining fairly tightly controlled survey conditions.

Another aspect of question ambiguity is the use of *vague words*; for example, the term “usual” conjures up different interpretations in different respondents. In general, it is advisable to avoid such terms, and ask instead about an actual event or activity for a specific day. Otherwise, the analyst will be put into the impossible position of trying to evaluate statistically what the “usual” event or behavior is. Also, when asking about activities during a certain time period, it is important to specify exactly the beginning and end of such a period (e.g., avoid using references like “last week”), since the beginning and end are subject to an individual’s interpretation.

Even survey designer neophytes will know not to use *loaded* questions (i.e., those that appear to solicit potentially embarrassing information, an example being “What was the purpose of your trip to the hotel?”). The appropriate strategy is to use filter questions in order to determine whether the next questions to be asked are actually relevant to the respondent.

Also to be avoided are *leading* questions. Leading questions are those that tend to influence the respondent (intentionally or not) in the direction of certain answers. For instance, if examples of possible answers are provided, the respondent will more likely draw an answer from the list of options provided, rather than bother to think of others that were not mentioned. Another example is the case where the question asked of the respondent implies that a certain answer will make the respondent appear in a better light. Richardson et al. (1) provide an extensive list of examples and discussions covering both “loaded” and “leading” questions.

The problem, however, is to recognize when a question has the undesirable attribute of being “loaded” or “leading”. Clearly, the analyst does not intend to formulate such questions, but it happens nevertheless. This is particularly true for leading questions, the answers to which will introduce serious bias, since they are likely to be consistent across respondents. There is a particular danger of leading respondents in interviewer-administered surveys, in the form of interviewer gestures and intonation, if the interviewer is not trained sufficiently to avoid such situations. In particular, when opinions and attitudes are solicited, violation of question neutrality can be devastating to the survey results.

A commonly accepted principle is to put more *sensitive questions* (e.g., about income) near the end of the instrument or interview, in the hopes that by that time, sufficient commitment to the survey topic has been generated to have increased the “comfort level” in the respondent to such a degree that it is easier and more natural to provide answers to those questions. One strategy that has been used to reduce nonresponse to the question of income in a household travel survey is to ask the question twice, once during recruitment, and then again in the household form sent to the respondent. Stopher and Metcalf used this strategy and created a new income variable by combining the responses from both questions when one of them was not answered. When two different responses were received, they used the second answer, on the assumption that this answer

was more reliable, since the respondent had more time to provide a more accurate answer. As can be seen in Table 1, the result of using this technique in two recent household travel surveys reduced the income nonresponse rate by 52%.

TABLE 1 Income Nonresponse Rates

	<i>Number of Cases</i>	<i>Percent of Households Refusing to Respond or Answering "Don't Know"</i>			
		<i>Recruit Income</i>	<i>HH Form Income</i>	<i>Combined Response</i>	<i>Improvement over HH Form</i>
Region 1	3,082	16.1%	8%	4.2%	52%
Region 2	16,084	22%	13%	6.8%	52%

Testing of Question Formulation and Instrument Design: The Case for Pretests

It is difficult to quantify the degree of inaccuracy and bias attributable to poor survey instrument design and faulty question formulation, due to the almost infinite number of "sins" that can be committed by the survey designer, and due to the great variety of survey purposes, designs, resource constraints, etc.

While all the guidelines referred to in this paper for instrument design and question formulation are useful and should be heeded, it is clear that there typically remains a significant difference between the good intentions of the survey designer and the effectiveness of the effort. The only known way of ascertaining whether the effort has met expectations is to perform a complete pretest that replicates the entire survey in all aspects and phases. This should be completed with a small sample from the actual population being surveyed; asking colleagues to complete the questionnaire is a very ineffective method of pretesting a survey. In 1996, NCHRP Synthesis 236 involved a "survey" of recent household travel surveys conducted in the United States, including questions about whether or not a pretest was conducted, and if so, which aspects of the survey were tested, and whether these items were changed in the main survey as a result of the pretest. The results indicated that 74% of the surveys included some kind of pretest, and 92% of those that were pretested were changed because of the pretest. While most of these surveys did not undergo a complete run-through of all aspects, Table 2 indicates that many pretests resulted in changes to items other than the questionnaire.

TABLE 2 Results of Survey Pretests

<i>Element Tested</i>	<i>Percent of Those Pretesting That Changed This Item After Pretest</i>
Instrument	85%
Management	65%
Training of Survey Personnel	60%
Sampling	31%
Key punching	N/A
Geocoding	100%
Analysis	18%
Incentive	11%

RECOMMENDATIONS AND FUTURE RESEARCH

A number of recommendations concerning survey design and question formulation can be extracted from this paper:

- Include legitimizing and motivational endorsement letter.
- Make objective of survey clear to the respondent.
- Design survey and questions from the respondent's perspective, not the analyst's.
- Provide clear survey and question instructions (where appropriate, by means of examples).
 - Keep in mind respondent's likely ability to provide answers.
 - Ensure professional design and layout of self-completion survey forms.
 - Provide logical and intuitively obvious sequencing of blocks of questions, and questions within such blocks.
 - Arrange questions from the more general to the more specific.
 - Obtain sociodemographic information first. In case of termination, at least inferences can be drawn about the characteristic of the nonrespondent.
 - Minimize writing requirements of respondent (other than in open questions).
 - Keep survey length to the minimum necessary. (Remember, the survey most likely is an imposition on the respondent.)
 - Perform reverse-order analysis with hypothetical data in the course of instrument design, to determine which questions are absolutely necessary to obtain the data required for the intended analysis.
 - Use simple language, appropriate for the target population. (This includes nonnative language forms and interviewers.)
 - Formulate questions in a totally unambiguous manner (lots of pretesting required!).
 - Set realistic limits for event recall period.
 - Provide memory jogger for travel diary surveys.
 - For interview surveys, perform extensive screening and training, as well as in-field supervision of interviewers.
 - In personal and telephone interview surveys, questions have to be posed verbatim, as dictated by the interview script, and they must be delivered by the interviewer without hint of suggested preference in expected answers.
 - In hypothetical questions (e.g., in stated preference surveys), ensure that the hypothesized situations described are reasonable and are likely to be within the respondent's mental grasp.
 - Use open questions sparingly, since the results tend to be difficult to use for quantitative analyses, and they usually considerably increase the cost of the survey.
 - Sensitive questions should appear near the end of the survey.

To advance the state of the art in survey design and question formulation, the best approach would be to engage in research surveys. This approach, while costly, would allow the scientific testing of alternative approaches to general survey design and question formulation, as was done, for example, by Socialdata, in connection with KONTIV (23.) Also, the meta-analyses referred to in this paper are producing promising avenues for

advancing the evaluation and understanding of cognitive processes involved in generating survey responses.

However, since research surveys are expensive, and furthermore, typically do not have immediate payoff for mission-oriented (transportation and planning) agencies, sponsorship tends to be difficult to obtain. Yet the long-term benefits in terms of obtaining data of greater validity and reliability, as a consequence of better response rates (quantity) and better quality (accuracy), would be significant.

In the meantime, the use of best-practice information should be increased substantially. A very significant body of knowledge and experience about travel surveys has been accumulated from around the globe over the last 15 to 20 years. And, as mentioned before, several explicit state-of-the-practice documents are available.

However, it can be observed time and again that the average transportation and planning agency staff member is frequently unaware of the many pitfalls of designing surveys and survey instruments, while remaining somewhat oblivious to the existence of helpful state-of-the-practice documents. A lot of publicity work remains to be done by the travel survey research community to bring current knowledge and experience to a wider audience.

ACKNOWLEDGMENT

The authors gratefully acknowledge the constructive comments provided by Mr. Ambrosius Baanders, Dr. Peter Jones, and Dr. Nelly Kalfs, as well as by anonymous reviewers, on the earlier drafts of this paper.

REFERENCES

1. Richardson, A.R., E.S. Ampt, and A.H. Meyburg. *Survey Methods for Transport Planning*, Eucalyptus Press, Melbourne, Australia, 1995.
2. Scherpenzeel, A.C., and W.E. Saris. "The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies." *Sociological Methods and Research* 25(3), 1997, pp. 341-384.
3. Andrews, F.M. "Construct Validity and Error Components in Survey Measures: A Structural Modeling Approach." *Public Opinion Quarterly* 48, 1984, pp. 409-422.
4. Rogers, W.L., F.M. Andrews, and A.R. Herzog. "Quality of Survey Measures: A Structural Modeling Approach." *Journal of Official Statistics* 8, 1992, pp. 251-275.
5. Költringer, R. "Measurement Quality in Austrian Personal Interview Surveys." In: *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, Chapter 12, edited by W.E. Saris and A. Münnich. Budapest, Hungary, Eötvös University Press, 1995.
6. Scherpenzeel, A.C., and W.E. Saris. "The Evaluation of Measurement Instruments by Meta-Analysis of Multitrait-Multimethod Studies." *Bulletin de Methodologie Sociologique* 39, 1993, pp. 3-19.
7. Alwin, D.F., and J.A. Krosnick. "The Reliability of Survey Attitude Measurement." *Sociological Methods and Research* 20(1), 1991, pp. 139-181.
8. Lord, F., and M.R. Novick. *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA, 1968.

9. Tourangeau, R., and K.A. Rasinski. "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Psychological Bulletin* 103, 1988, pp. 299-314.
10. Schwarz, N., F. Strack, H.J. Hippler, and G. Bishop. "The Impact of Administration Mode on Response Effects in Survey Measurement." *Applied Cognitive Psychology* 5, 1981, pp. 193-212.
11. Schuman, H., and S. Presser. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. New York, Academic Press, 1981.
12. Saris, W.E., and F.M. Andrews. "Evaluation of Measurement Instruments Using a Structural Modeling Approach." In: P.R. Biemer et al. (eds.), *Measurement Errors in Surveys*, Wiley Interscience, New York, 1991, Chapter 28, pp. 574-597.
13. U.S. Department of Transportation and U.S. Environmental Protection Agency. "Travel Survey Manual" (Travel Model Improvement Program), prepared by Cambridge Systematics, Inc., Washington, D.C., July 1996.
14. National Opinion Research Corporation (NORC). "Reducing Nonresponse." Unpublished preliminary draft chapter of report on travel survey nonresponse problems. Sponsored by US DOT, FHWA, Office Highway Information Management, March 1997.
15. Heberlein, T.A., and R. Baumgartner. "Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature." *American Sociological Review* 43, 1978, pp. 447-462.
16. Yammarino, F.J., S.J. Skinner, and T.L. Childers. "Understanding mail survey response behavior." *Public Opinion Quarterly* 55, 1991, pp. 613-639.
17. Oppenheim, A.N. *Questionnaire Design, Interviewing and Attitude Measurement*. Pinter Publishers, London, 1992.
18. Stopher, P.R., L. Shillito, D.T. Grober, and H.M.A. Stopher. "On-Board Bus Surveys: No Questions Asked." *Transportation Research Record 1085*, National Research Council, Washington, D.C., 1986, pp. 50-57.
19. Clarke, M.I., M.C. Dix, and P.M. Jones. "Error and Uncertainty in Travel Surveys." *Transportation* 10, 2, 1981, pp. 105-126.
20. Meyburg, A.H., and W. Brög. "Validity Problems in Empirical Analyses of Non-Home Activity Patterns." *Transportation Research Record 807*, National Research Council, Washington, D.C., 1981, pp. 46-51.
21. Stopher, P.R., and Sheskin. "A Method for Determining and Reducing Nonresponse Bias." *Transportation Research Record 886*, National Research Council, Washington, D.C., 1982, pp. 35-41.
22. Bradley, M., and A. Daly. "Use of the Logit Scaling Approach to Test for Rank-Order and Fatigue Effects in Stated Preference Data." *Transportation* 21, 2, 1994, pp. 167-184.
23. Brög, W., K. Fallast, H. Katteler, G. Sammer, and B. Schwertner. "Selected Results of a Standardized Survey Instrument for Large-Scale Travel Surveys in Several European Countries." In E.S. Ampt, A.J. Richardson, and W. Brög (eds.), *New Survey Methods in Transport*. VNU Science Press: Utrecht, The Netherlands, 1985, pp. 173-192.
24. *NCHRP Synthesis of Highway Practice 236*. "Methods for Household Travel Surveys" (prepared by Peter R. Stopher and Helen M.A. Metcalf). Transportation Research Board, National Research Council, Washington, D.C., 1996.

25. Stecher, C.C., S. Bricka, and L. Goldenberg. "Travel Behavior Survey Data Collection Instruments." *Conference Proceedings 10*, Conference on Household Travel Surveys: New Concepts and Research Needs (Irvine, CA, March 12-15, 1995). Transportation Research Board, National Research Council, Washington, D.C., 1996, pp. 154-169.
26. Bonsall, P., M. Wardham, C. Nash, and P. Hopkinson. "Development of a Survey Instrument to Measure Subjective Valuations of Non-Use Benefits of Local Public Transport Services." In: E.S. Ampt, A.R. Richardson, and A.H. Meyburg (eds.), *Selected Readings in Transport Survey Methodology*. Proceedings of the 3rd International Conference on Survey Methods in Transportation (Washington, D.C., January 5-7, 1990). Eucalyptus Press, Melbourne, Australia, 1992, pp. 71-87.
27. Brög, W., A.H. Meyburg, and M.J. Wermuth. "Development of Survey Instruments Suitable for Determining Non-Home Activity Patterns." *Transportation Research Record 944*, National Research Council, Washington, D.C., 1983, pp. 1-12.

Workshop Summary

Ambrosius Baanders

*Transport Research Centre AVV
Netherlands Ministry of Transport, Public Works and Water Management*

Stephen Arrowsmith

Statistics Canada, Ottawa, Ontario, Canada

INTRODUCTION

The act of asking respondents about their attitudes and behavior can influence both of these; as in physics, the observer can influence the subject of the research. In addition, the answers obtained can be influenced by the precise wording of the questions, and the sequencing and general context in which they are asked. The medium of communication can have a more general influence on the quality of response. The aim of workshop 12 on questionnaire design was to identify and address these issues and make recommendations on best practice, giving an indication, where possible, of the size and direction of any biases, and advice on how to reduce such problems.

The workshop discussion started from the resource paper by A. Meyburg (1997), and its presentation by the discussant H. Metcalf. The participants felt that many of the techniques that should be implemented to avoid problems when designing a questionnaire are well-known in the transport research profession and are well-documented in the literature. The resource paper lists many of those. But the problem, it was felt, is that there is a big difference between the state of the art and the state of the practice. The problem with the state of the practice is indicated in the preface of Fowler's book, *Improving Survey Questions: Design and Evaluation*, when he opens by stating:

I believe that the design and evaluation of survey questions is the most fertile current methodological area for improving survey research. Poor question design is pervasive, and improving question design is one of the easiest, most cost-effective steps that can be taken to improve the quality of survey data (1995, p. vii).

To assist in efforts to enhance the state of the practice, the workshop produced a number of guidelines for designing transport survey questionnaires. In its discussions, the workshop touched on a number of subjects where it was felt that more research is needed. In this way, it produced a research agenda. Finally, the workshop came up with a number of recommendations on how to organise the further development of research on questionnaire design.

The workshop noted that many different types of questionnaires are found in transport research. They may differ by the way the respondent is approached, be it a personal interview, a telephone interview or a written (self-completion) questionnaire, handed out or sent through the mail. The written questionnaires may either be mailed back or reviewed in a later telephone interview. The respondent may be interviewed at home or be intercepted during a trip. It is also possible that commercial enterprises and their employees, such as freight shippers or truck drivers, could be respondents. In sum, almost

anyone could be a respondent to a transportation survey, contingent on the objectives of the survey.

This report deals separately with each of the items in terms of potential guidelines, future research and organizational recommendations. Within each of the item components, there is no delineation made by questionnaire type, unless there is a requirement for separation. We have assumed that many of the criteria are the same, as was the case during the discussions.

GUIDELINES

The guidelines identified below reflect the items that were discussed during the workshop. As workshop discussions tend to move around, the items are not listed in any particular sequence.

PREPARATION TIME, TESTS, AND PILOTS

The big difference between the state of the art and the state of the practice is caused to a large extent by client ignorance, poor survey practices, timelines and cost considerations.

Many clients of transport surveys are unaware of the pitfalls of questionnaire design, and, often, the time between the decision by the client to go ahead and the actual start of the interviewing phase is too short to properly design the survey instruments. This may be because the time at which the respondents have to be interviewed is constrained, e.g., by the wish to use a “shoulder period” in order to exclude unwanted seasonal effects. In light of this problem, the workshop felt that the questionnaire designer should convince the client that it is absolutely necessary to think about the design, and to undertake cognitive testing, some combination of field-testing, and/or a full “dress rehearsal” pilot. This process will enable the careful checking of the design, and allow time required to incorporate the ensuing modifications. Although this may sound trivial, the workshop participants knew of many instances where this was not done properly, resulting in downstream data problems either prior to or during the analysis phase of the project. This point cannot be overemphasized.

While there are many important elements to be tested in pretests, the ability and accuracy with which the respondents are able to provide answers to questions may be the most important. If people cannot provide an answer, cannot provide an accurate answer, or have various interpretations of a question, the data are of little or no use. It is always critical to ask, do we know what we are measuring, and are we measuring what we think we are measuring?

In the case of trips, it is possible that the feasible recall period increases when the trip in question is made less frequently (becomes more of a special event in the perception of the respondent). For daily mobility, the recall period is very short, and therefore, the use of trip diaries/logs is recommended.

It was pointed out that, in some cases, the above considerations may mean that the designer will have to say “no” to the client, because there is not enough time to prepare the survey properly (there may also be ethical reasons to do this). This may be very difficult when the survey practitioner is in a competitive situation, but it may be necessary to convince the client that proper preparation is absolutely necessary. The old adage “garbage in - garbage out” is also true here.

Language

Some societies have always known a mix of languages, while in other societies, the coexistence of different languages is a fairly recent phenomenon, e.g., through immigration. But even in a country where only one language is spoken, the problem of different languages will arise in such surveys as those monitoring international air travellers. Therefore, it is incumbent on the questionnaire design team to take into account language issues that may be pertinent to the survey.

The nature of languages is such that the problem of language should be handled very carefully. A translation that is correct in a general context does not always convey the same meaning when used in describing travel behavior. An example of how complicated this can be was given by P. van der Reis, describing travel questionnaires in South Africa (1997). The English word “convenient”, which translates into “gerieflik” in Afrikaans and into “lungile” in Xhosa, was found to have a totally different meaning for most respondents in each of the three languages when used in the context of travel to and from work (“close to work”, “close to home” and “uncrowded”, respectively). A good way to check for such problems is to have the questionnaire text translated back into the original language, to see if it is the same.

In a number of cases, pictorials are used to overcome the problems of having different languages in the group to be surveyed. However, even in these cases, verification is essential to ensure that the pictorials convey the intended meaning. In another example from South Africa, it was noted that certain African cultures understand pictures in a way totally different from European cultures.

Another important element in questionnaire wording is to use the language of the respondent. In this context, “language” should be taken in a broad sense. With any language, there are groups of people who think and speak in a different way, and who attach different meanings to the same words. The “language” of truck drivers is an example of this, and in a survey of truck drivers, the peculiarities of this “language” should be respected. In the same way, one should avoid the typical jargon used in transport research, which will not be understood by the average member of the public, given that most people are not very familiar with transport issues.

Reverse Analysis and Respondent Burden

A common experience in the workshop was that, often, some of the questions asked during the survey were not used later in the analysis. If you know in advance that a certain question can or will not be used in the analysis (due to low levels of incidence, hence a lack of reliability, or because, at the end of the day, the variable in question is just not that important), then the question should not be asked. By not asking the question, respondent burden is reduced. A way to do this, as was suggested in the resource paper, is by first doing a “reverse analysis”, i.e., do the analysis first with imaginary data. This will show which variables are really needed, and only these should be used when designing the questionnaire. Another approach is to have those doing the analysis create a set of table shells (based on a set of hypotheses) to determine where data are required. This, of course, takes time at the preparation stage of the survey, but ultimately saves time later.

Participants stressed that a good questionnaire design should minimise the burden placed on the respondent in every respect. The length of the questionnaire in terms of time

to complete, the thickness of the document in terms of the perceived burden for self-completed forms, and the complexity of the questions in terms of difficulty to complete are all important considerations. But as workshop 6 dealt specifically with respondent burden (Muranami, 1997), the other aspects of burden were not discussed further.

Questionnaire Typography and Layout

The use of clearly readable letters (font size and font type) in written questionnaires is very important in the experience of the participants. In the case of the Latin alphabet, questions should be typed in lowercase letters, because that makes reading easier. The ease of reading should be tested in the pilots.

Elaborate instructions make a questionnaire (especially a written one) look more complicated. They should be kept to a minimum. Definitions and explanations of what exactly is meant should be incorporated into the question itself as much as possible, or should precede the question.

As geocoding is important in many surveys, sufficient space should be provided on the forms to write all the necessary details; this is especially the case for self-completed questionnaires. Items such as street name and number, building or landmark should be asked for explicitly. The name of the municipality needs extra care in the case of large urban areas, where the municipal boundaries are often not obvious to many people.

Using recycled paper for the written questionnaires, and indicating that by use of the “recycling” logo, was found to increase the response rate in some countries.

A Research Agenda

As mentioned previously, the workshop participants felt that many clients of transport surveys are unaware of the pitfalls of questionnaire design, and the relationship to final output. Moreover, their concerns are immediate and, typically, problem-centered. Consequently, they are typically not willing to fund research on questionnaire design and development as part of the research projects they commission. As a consequence, innovation in questionnaire design was found to be slow.

In the course of the workshop discussion, there were points where participants felt there was insufficient knowledge to formulate a guideline, and it was therefore suggested that the item in question be considered for further research. Identified below are the items, which could constitute a kind of research agenda:

Improving the Effectiveness of Questionnaires

Improving the effectiveness of questionnaires was thought to be possible by:

- Finding ways to reduce repetition in both written and oral questionnaires (e.g., when respondents are traveling together, or when a respondent makes the same trip every day during a multiday survey);
- Developing pictograms to be used instead of words in self-completion questionnaires where literacy or language differences are a problem (but the understanding of the pictograms should be checked);

- Increasing our understanding of the effects of layout and colour in the forms used for self-completion questionnaires;
- Testing the differences in quality and response rates through an examination of the effects of sending one long questionnaire or holding one long oral interview (face-to-face or by telephone) versus multiple shorter questionnaires or interviews;
- Finding ways to overcome the problem of the underreporting of trips (paying special attention to the propensity of respondents, under some circumstances, to report that they stayed at home when they did make some trips);
- Finding ways to reduce the impact of normative behavior on respondent answers;
- Undertaking research into the recall periods for long distance trips (which are not frequent enough to be surveyed using a trip diary);
- Finding ways to overcome the dilemma between open-ended responses and precoded response categories; identifying the types of questions where one approach is better than the other;
- Conducting more tests on the use of instructions for the respondents, both in self-completion and in oral questionnaires; and
- Undertaking more comparison tests between asking all questions through self-completion forms sent by mail, or using a hybrid approach where some of the questions are asked by telephone and other questions are asked on a mailed self-completed form.

With regard to self-completion trip diaries, the discussion was whether columns or rows should be used for the different attributes that have to be recorded for each trip. Another question was whether a trip diary should preferably be administered in booklet form, or as separate sheets that form a log upon completion.

The dilemma between open-ended responses and precoded responses is that the latter require less thinking from the respondent and thus are easier to give, but they direct the respondent toward the preestablished alternatives. Open-ended responses are more likely to generate answers that the designer did not think of beforehand, thus giving more insight into the respondent's behavior. However, the coding of answers to open questions requires more skill from the person doing the coding. Another problem with precoded answers is the choice of the number of answer alternatives. If the number is small, the respondents may not find an answer that applies to their situation. If the number of choices is large, the respondents may find it difficult to make a choice. In any case, the answer alternatives should be worded very carefully to ensure that categories are mutually exclusive. These problems are particularly relevant for self-completion questionnaires.

In the case of oral interviews, a possible solution is that the interviewer poses an open question, and interprets the answer by choosing from a set of precoded answers. This requires special skill from the interviewer, but with proper training, this is not considered to be an insurmountable problem. The issue of combining precoded and open-ended response categories was discussed by the workshop, with thoughts directed toward the creation of a decision-tree approach to asking the questions. This is an area for further research, both conceptually and operationally.

Trip Stage Information and the Use of GPS

More research is needed, it was felt, to find the best way to collect data on the different stages of a trip or journey (e.g., for time information, how much rounding off should be allowed?), and to distinguish changes of mode. In this context, experiments using geo-positioning systems (GPS) were noted. This was seen as one possible way forward, but other methods should also be examined.

Better definitions are needed for “trip” and “mode”, to describe a walking trip, and to describe working from home (teleworking).

The Linking Between CATI and GIS

For oral interviews, computer-aided interview techniques (CATI, etc.) were an innovation well-appreciated by the participants. It was suggested that by linking CATI software with geographic information systems (GIS), the checking for mistakes in addresses given by respondents could be much improved in some situations.

Use the Knowledge from Other Behavioral Sciences

Despite the fact that interviewing of respondents is done in many behavioral sciences, the participants felt that transport analysts do not sufficiently use the experience from other fields. It was recommended that a thorough literature review be conducted into the relevant experience from other fields, notably, cognitive psychology.

SOME RECOMMENDATIONS

The workshop came up with two recommendations, which require the initiative to be taken by an overarching organisation. It is hoped that the organization that was recommended by Workshop 8, on practitioner’s needs (Liss, 1997), will be able to take such initiatives. The recommendations are:

- Create a system or platform for the exchange of research methodology and experience between those involved in questionnaire design.
- Develop a standardization of concepts for use in transport survey questionnaires.
- Develop a set of ISO norms for this.

REFERENCE

1. Floyd J. Fowler. Improving survey questions: design and evaluation. *Applied Social Research Methods Series*, Vol. 38. Sage Publications, Thousand Oaks, Calif., 1995.