SAGE researchmethods

Questionnaire Design

In: Encyclopedia of Survey Research Methods

By: Gregory G Holyk
Edited by: Paul J. Lavrakas
Book Title: Encyclopedia of Survey Research Methods
Chapter Title: "Questionnaire Design"
Pub. Date: 2011
Access Date: June 3, 2019
Publishing Company: Sage Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412918084
Online ISBN: 9781412963947
DOI: https://dx.doi.org/10.4135/9781412963947
Print pages: 657-659
© 2008 Sage Publications, Inc. All Rights Reserved.
This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

2008 SAGE Publications, Ltd. All Rights Reserved.

Questionnaire design is the process of designing the format and questions in the survey instrument that will be used to collect data about a particular phenomenon. In designing a questionnaire, all the various stages of survey design and implementation should be considered. These include the following nine elements: (1) determination of goals, objectives, and research questions; (2) definition of key concepts; (3) generation of hypotheses and proposed relationships; (4) choice of survey mode (mail, telephone, face-to-face, Internet); (5) question construction; (6) sampling; (7) questionnaire administration and data collection; (8) data summarization and analysis; (9) conclusions and communication of results.

One goal of the questionnaire design process is to reduce the total amount of measurement error in a questionnaire. Survey measurement error results from two sources: variance and bias. Question wording (including response alternatives), ordering, formatting, or all three, may introduce variance and bias into measurement, which affects the reliability and validity of the data and conclusions reached. Increased variance involves random increases in distance between the reported survey value from the "true" survey value (e.g. attitude, behavior, factual knowledge). Increased bias involves a systematic (directional) difference from the "true" survey value. Questionnaire-related error may be introduced by four different sources: the interviewer, the item wording/format/ ordering, the respondent, and the mode of data collection. Careful questionnaire design can reduce the likelihood of error from each of these sources.

Coals, Objectives, And Hypotheses

The process of designing a questionnaire begins with a definition of the goals and objectives of the study. A clearly defined purpose acts as an anchor that sets the stage for questionnaire format and question construction or ordering. In what concepts and phenomena is the researcher interested? What does he or she want to learn? To what population are the results to generalize? These are all questions that should be asked during the initial phase of questionnaire design. Each item included in the questionnaire should meet the criteria that it will provide useful information related to the goals and research questions of the survey; parsimony is a virtue in questionnaire design. To properly define goals and research questions, key concepts of interest need to be detailed. A list of concepts of interest and how they relate to one another aids in selecting specific questions to include. These concepts are transformed into (i.e. operationalize as) survey questions that measure the concept in some way, proceeding from abstract concepts to specifie measurements. Once the goals and objectives of the survey have been established, it is possible to generate specifie research questions and, possibly, directional hypotheses based on theory and previous findings. The items used in the questionnaire should produce data that are appropriate to the desired analyses.

Questionnaire Format

The layout of a questionnaire, no matter what type, should reduce the cognitive burden of respondents (and interviewers) and contain an intuitive and logical flow. For example, in most cases, questions on related

2008 SAGE Publications, Ltd. All Rights Reserved.

topics should be grouped together and questions should maintain the chronology of events. Questionnaire format should be as easy as possible to understand and use for the interviewer or the respondent depending on the mode of administration. Questions should be numbered individually, clearly spaced, and visually distinct from one another. It is especially important that self-administered questionnaires provide clear and concise instructions and have a simple layout (e.g. common question format, visual instructions for skipping questions). For interviewer-administered questionnaires, instructions that are to be read aloud to respondents should be visually distinct from instructions that are for only the interviewer; for example, set off by italics, CAPS, bold type, or parentheses ( ). Ideally, important questions should appear early in a questionnaire to avoid the possible negative effects of respondent fatigue on motivation, recall, and item nonresponse.

Questionnaires that have a professional appearance are taken more seriously by respondents (e.g. displaying professional affiliations). Social validation is also an important factor; questionnaires that end by thanking respondents for their time and effort and reminding respondents of the usefulness of the data they have provided are rated as more enjoyable by respondents.

Questionnaire length is an important issue for both cost reasons and effects on respondent behavior. First, longer questionnaires place a greater burden on respondents and are more costly. Second, longer questionnaires can have many significant effects on respondent behavior. Refusal rates rise with the length of the questionnaire; respondents may be willing to spend time on a 20-minute interview but may feel a 45-minute interview is far too long. In addition, greater attrition, measurement error, and satisficing may occur due to fatigue with longer questionnaires.

The contingencies that are built into a questionnaire, even when the questionnaire is computer assisted, must be designed with impeccable logic. Furthermore, when the questionnaire is not administered via computer and is self-administered (i.e. paper-and-pencil questionnaires), these so-called skip patterns must be explained in very clear and uncomplicated language, graphics, or both.

Question Wording

Unclear concepts, poorly worded questions, and difficult or unclear response choices may make the questionnaire difficult for both respondents and interviewers. Questionnaires should contain items that are both reliable and valid. Reliability is the consistency of the measurement; that is, the question is interpreted and responded to similarly over repeated trials. Construct validity is whether or not the measurement, as worded, properly reflects the underlying construct of interest.

Questions may be troublesome for respondents with respect to comprehension, retrieval, and response formation. First, the respondent must be able to understand the question. Questions should use simple (enough) terms and concepts that respondents are likely to know, and they should not contain vague and ambiguous terms; in other words, questions should be worded appropriately for the reading and knowledge level of the respondents. Questions should avoid unnecessary complexity, such as compound sentences,

2008 SAGE Publications, Ltd. All Rights Reserved.

double-barreled questions (two questions within one), and double negatives. Longer questions that provide sufficient detail (without becoming convoluted) and are explicit generally will enhance comprehension.

Even though questions may be perfectly understandable for respondents, there may be problems with retrieval of the desired information from respondents' memory. Questions should ask for information that is as recent as possible and provide cues that match the original encoding context. Question construction may be either personal or impersonal in form: The choice depends on what kind of information, personal or general, the question is designed to elicit. Shorter recall periods (or reference periods) are easier for respondents to handle. Cues that are distinctive and personally relevant to the respondent also aid retrieval. Questions that ask for information that is general or was initially superficially encoded in the brain are prone to greater error in comparison to questions with appropriate, specific cues.

Terms that have multiple possible interpretations should have brief definitions provided in the question. Questions that define concepts for respondents work better if the definition of the concept precedes the actual question. Likewise, reference periods are also more effective when they are provided before the question. Event history calendars are useful for asking about similar events over a long period of time. It is also a good idea to avoid sensitive questions if possible (e.g. questions about income, religion, or sexual behavior). Any questions relating to sensitive information should be clearly necessary, and respondents should understand this need.

Response Alternatives

Responses can be either open-ended or closed-ended. Open-ended questions provide no predetermined response categories and allow the respondent to answer with whatever information he or she considers relevant. This answer format is more cognitively demanding for respondents, but it often results in more detailed and informative responses. Open-ended questions are good for gathering information on a topic for which there is no clear set of response categories. Open-ended questions work better with interviewer-administered questionnaires than in surveys that use self-administered questionnaires. An open-ended response format may result in a great deal of information, but it may be information that is not easily comparable or easily codable. Open-ended data also can be quite time-consuming and very costly to process.

Closed-ended questions ask respondents to select among a predetermined set of response categories. These response categories must be exhaustive and mutually exclusive. The closed-ended method reduces the cognitive burden of the respondent and enhances the ability to compare responses. The data are already coded (assigned a numerical value) and can be easily quantified, which saves data processing time and money. Closed-ended questions are ideal for self-administered questionnaires because they avoid the greater subjectivity and volatility of open-ended questions. However, if the researcher is not careful, the selection of response alternatives may bias respondents by framing thinking and by predetermining what are considered "appropriate" answers.

2008 SAGE Publications, Ltd. All Rights Reserved.

Response categories for event information may be divided by occurrence, absolute frequency (actual amount), relative frequency (quantifiers that denote more or less of something, such as "always," "most of the time," "sometimes," "basically never"), regularity, and specific dates. Attitude answers may take the form of ratings, rankings, agree-disagree scales, and forced choice. Attitude questions essentially require an evaluation of some phenomenon (e.g. approval/disapproval of the president). In selecting terms to use to describe attitudes, it is best to use terms that are simple and easy to understand, as well as terms that are likely to be interpreted similarly by all respondents.

Response scales require decisions concerning the length of scale, whether or not to include midpoint or explicit "don't know" options, and whether to use verbal labels, numbers, or both. A very common scale for closed-ended survey items is a Likert scale (e.g. "strongly agree," "somewhat agree," "neither agree nor disagree," "somewhat disagree," or "strongly disagree"). Rating scales can be presented as either a verbal description (e.g. strongly agree, strongly disagree), numbers (e.g. 1 to 5), or a combination of both. Meaningful verbal labels with relatively equal intervals provide a linguistic reference point for numerical scales, which significantly aids respondents.

Including a middle category is desirable when there is a readily identifiable and meaningful midpoint to a scale (e.g. increase, remain the same, decrease). The number of response options should attempt to maximize discrimination between response categories while maintaining the usefulness and meaningfulness of response categories. Research shows that five to nine response categories is best. There is tremendous debate over the usefulness of explicitly including "don't know" responses. Some argue that questionnaire designers are incorrect in assuming respondents always have an opinion and that surveys often force respondents to artificially produce an opinion. The lack of opinion regarding a topic has been referred to as displaying a "nonattitude." Others argue that explicit "don't know" response choices encourage respondents to satisfice and thereby not make the mental effort required to answer questions accurately. If given the chance, respondents with a low level of motivation will opt to get through a survey quickly and often choose the "don't know" response to facilitate this process, when otherwise they would have produced an opinion by not being offered a "don't know" option and thus been forced to put a little more effort into answering the question. A good situation in which to include an explicit "don't know" option is on unfamiliar topics where respondents are much more likely to have no knowledge or information about the subject and thus truly may have no opinion.

Question Testing Methods

Survey methodologists have developed several different methods of testing questionnaires to detect problems before they go out in the field. Question testing methods include the following: expert panels, traditional interviewer debriefing, behavior coding, cognitive interviewing, focus groups, and split-half experiments. Using one or some combination of methods improves the quality of a questionnaire.

Conventional pretesting involves interviewers conducting a relatively small number of interviews with a draft

2008 SAGE Publications, Ltd. All Rights Reserved.

version of the questionnaire, followed up by an informal group debriefing session. Conventional pretesting accurately identifies problems that interviewers may have with the delivery of questions, but it is not as good at diagnosing respondent difficulties.

Behavioral coding standardizes the method of observing and recording the respondent-interviewer interaction. Behaviors that occur during the question-answer process are categorized (i.e. respondent asks for repetition of the response categories), and their frequency of occurrence is recorded. Behavioral coding aids in the identification of problems in question wording and question ordering and conceptual ambiguities. This method generates more reliability and consistency across interviewers in coding the question answering process because it uses a method that results in quantifiable data. One drawback to behavioral coding is that it does not yield any information for the reasons behind question problems.

Cognitive interviewing techniques are adept at revealing the largely invisible cognitive processes respondents engage in while answering questionnaires. Respondents may be asked to "think aloud" while answering questions, or respondents may be asked specific standardized question probes (e.g. "What did you count as ‘exercise’ when I asked you, On average how many times do you exercise per week?") designed to reveal the thinking process. This method is useful for identifying problems with question comprehension, recall strategies, sensitivity of questions, response selection, and vagueness of reference periods. The most common criticism of cognitive interviewing techniques is that they are somewhat artificial, sometimes causing the respondent to engage in behavior and thinking patterns they would not normally engage in when responding to a question.

Expert panels gather substantive and methodological experts to critique questionnaires. These panels are a relatively inexpensive and quick alternative to the other question testing methods. Experts have "seen it all" over the years and can easily point out troublesome aspects of questionnaires.

Focus groups, that is, small groups of respondents asked to discuss survey questions, are able to generate a large amount of information in a short amount of time. However, this technique sometimes may result in a skewed group consensus due to group dynamics.

Split-half experiments examine the effects of different question wording, ordering, or format by randomly assigning different forms of a question to different respondents. However, these tests can be costly and may require larger sample sizes than a survey otherwise needs.

Gregory G Holyk

http://dx.doi.org/10.4135/9781412963947.n425

See also

- Attitude Measurement
- Behavior Coding
- Bias

2008 SAGE Publications, Ltd. All Rights Reserved.

- Closed-Ended Question
- Cognitive Interviewing
- Construct Validity
- Contingency Question
- Event History Calendars
- Exhaustive
- Focus Group
- Forced Choice
- Interviewer Debriefing
- Mutually Exclusive
- Nonattitude
- Open-Ended Question
- Paper-and-Pencil Interviewing (PAPI)
- Questionnaire
- Questionnaire Length
- Questionnaire-Related Error
- Question Wording as Discourse Indicators
- Reliability
- Respondent Burden
- Response Alternatives
- Satisficing
- Validity
- Variance

Further Readings

Aday, L. A.(2006). Designing and conducting health surveys: A comprehensive guide (3rd ed.). San Francisco: Jossey-Bass.

Dillman, D. A.(2006). Mail and Internet surveys: The tailored design method (2nd ed.). New York: Wiley.

Fowler, F. J., Jr.(1993). Survey research methods (2nd ed.). Newbury Park, CA: Sage.

Fowler, F. J., Jr.(1995). Improving survey questions: Design and evaluation. Thousand Oaks, CA: Sage.

Oppenheim, A. N.(1999). Questionnaire design, interviewing, and attitude measurement. New York: Pinter.

SchaefferN.C.PresserS. The science of asking questions. *Annual Review of Sociology*29(2003)65–88.