

1 The standardized survey interview

1 Introduction

Social science research is primarily devoted to describing and analyzing peoples' actions and attitudes. In modern states, governments need this information in order to make well-founded policy interventions and evaluate their effectiveness. For example, governments may want to know parents' attitudes towards the schooling of their children and whether or not proposed changes will meet with approval. Also, political parties need to know what issues their prospective voters consider important in order to adapt the election strategies accordingly. In addition, industries may want to know what the trends in fashion will be in two years time, in order to purchase the appropriate materials and dyes.

To gather such information, researchers may send people a questionnaire to fill out and return. Another increasingly popular procedure is to interview people, either in person or, far more economically, by telephone. In the 1980s, personal computers led to the introduction and rapid development of computer-assisted telephone interviewing (CATI).

Among the different types of research interviews, the standardized survey interview is the most prevalent one in use in western societies. Large numbers of people earn an income in the survey research industry, enormous amounts of public and private money are spent in this field, and many important decisions made by governments and companies are based on the results of survey research. In fact our lives are ruled by survey research to a considerable extent.

The standardized survey interview is designed for gathering data with which to measure the intentions, actions, and attitudes of large numbers of people, usually representative samples of the populations being studied. Measurement must be carried out in such a way that measures can be aggregated to describe the population. According to Fowler and Mangione, "The standard for the success of a survey is how well the data measures the aspects of the population the researchers are trying to describe. The goal of a survey is to produce accurate statistics" (1990:

2 Interaction and the standardized survey interview

12). The quality of research methods depends both on validity and reliability. The research method is valid when it generates the research data that it is designed to collect. This means that the questions should measure the dimension or construct of interest (construct validity). In addition, the respondents should interpret the questions as the researcher intended, and the answers should correspond with respondents' actual perceptions (response validity). So, when we want to know what percentage of the research population visited a museum last year, we should make sure to define exactly what we mean by "museum" (do art galleries count as museums?) and by "visiting" a museum (does a visit count if we briefly drop in, or if we only visit when we accompany our grandmother?). Furthermore, the research method must be reliable. The concept of reliability refers to the degree of variation among responses in repeated trials and with different interviewers. All respondents should understand the concept of museum in the same way, irrespective of the conditions under which they were interviewed.

2 Measurement error

Survey measurement is carried out by applying the same procedure to each respondent, so differences between responses can be attributed to real differences in the studied population. Differences in measures should not be attributable to characteristics of either the measurement instrument or measurement procedure. In order to reduce such measurement errors, the survey interview must be maximally standardized across all aspects of the measurement process.

- The interview setting, for instance either in person or over the telephone, should be the same for all respondents in order to avoid mode effects. For an overview of these so-called mode effects, see Lyberg and Kasprzyk (1991) and Czaja and Blair (1996).
- Characteristics of the interviewers should not influence the respondents' behavior. Research shows that interviewer characteristics such as sex, race, age, level of experience, or motivation can have an effect on response behavior (Schuman and Presser 1981; Fowler and Mangione 1990).
- All respondents must be presented with the same stimuli. Thus, every interviewer should present the questions exactly as scripted in the questionnaire, without omitting or adding anything.
- Interviewers should probe answers in a neutral way. They must follow the rules of standardized interviewing behavior in order to reduce interviewer-related errors.

The need for standardized interviewer behavior is partly related to the careful design of the questionnaire. Ideally, questionnaires are designed so that they take into account the various insights that survey methodology research has generated over the years. Since questionnaires are, ideally, designed so that maximum validity is guaranteed, interviewers should not decrease this.

Standardized interviewer behavior is also necessary for reasons of reliability. Survey methodology works from a stimulus-response model of interaction (Foddy 1993), in that every respondent should be presented with the same verbal input. For that reason, interviewers should read aloud each question exactly as worded, preserve the order of the questions, and probe in a non-leading manner. This consistency ensures that differences in responses can be related to respondent characteristics, rather than to characteristics of the interviewing procedure.

In section 3, I will briefly discuss how question format, wording, and order, as well as the order in which response options are presented, may have an effect on response behavior. For a more detailed discussion, see Schuman and Presser (1981), Belson (1981), Biemer et al. (1991), and Schwarz and Sudman (1992). At the end of this section I will make some remarks on the assumption that standardized question wording equals standardized meaning, even though the studies under review point in a different direction. Next I will focus on aspects of standardized interviewing techniques (section 4), and how to probe inadequate answers (section 5). In section 6, I propose that the concepts of interviewer effect and interviewer error are misleading and unfair. In fact, many supposed interviewer effects and errors should be seen rather as questionnaire effects and errors. Questionnaires, I claim, often do not live up to their intended use in interaction. Section 6 also points out why survey methodology may benefit from conversation analysis studies on survey interviews.

3 Question characteristics

3.1 *Question format*

Survey methodology distinguishes between different question formats. An important distinction is whether or not respondents are free to formulate their own answers. Question formats also differ depending upon who will do the final coding of the respondents' answers.

Open questions do not limit response alternatives. Respondents are free to formulate their own answers. The interviewers are instructed to copy the answers verbatim. Later in the research process, trained coders

4 Interaction and the standardized survey interview

will translate the copied answers into a restricted number of response codes.

In the case of closed questions, also known as fixed-choice questions, the interviewers read the questions and a list of response options from which respondents are to choose their answers. When respondents formulate their answers in a way that differs from the response options, interviewers are supposed to probe for a formatted answer (Schaeffer and Maynard 1996),¹ that is, an answer that matches one of the pre-coded response options that can be recorded. In these cases, it is crucial that interviewers use neutral probes. A special type of closed question is the yes–no question. Unlike ordinary conversation (Churchill 1978), the yes–no question constrains respondents replies, permitting only “yes” or “no” answers.

Survey questionnaires may also contain field-coded questions. These questions are presented to respondents as open questions. However, the interviewers are given a list of response options, and they must translate the respondent’s answer into a coded option before checking one of the response boxes.

Face-to-face interviews provide the possibility of using show card questions. Response options are written on a card that is presented to the respondents. After hearing the question, they select the pre-coded response of their choice. Whether or not the response options are presented on a card, closed questions require that respondents provide the information they have in mind in terms of the pre-coded response categories. Thus respondents are doing the coding work.

3.2 *Question-order and context effects*

The way a question is interpreted may depend on what precedes it. This holds true especially when successive questions concern the same topic. When the prior question provides a context for interpretation of the next one, we may speak of “question-order effect” or “context effect.” For example, if a person is asked how his wife is doing and is next asked how his family is doing, this person may not include his wife in the second answer. He could assume that the questioner would not request information already provided.

Context effects may also arise when respondents try to appear consistent by putting a response to a later question in line with an earlier one (the part–part consistency effect). See Schuman and Presser (1981) for an extensive discussion of several types of question-order effects. See also Schwarz and Sudman (1992).

3.3 *Response options and response order*

Closed survey questions present two or more response options from which the respondent is supposed to choose. Several experiments show that the presenting order of these options may influence the respondent's choice. This type of context effect is referred to as the "response-order effect." Respondents who read the response options themselves will be more likely to choose the options they see first. When options are read to them, however, respondents are more likely to endorse the ones they hear last (Krosnick and Alwin 1987). In order to reduce response-order effects, computerized questionnaires may randomize the order in which the response options are presented across respondents (Schwarz and Hippler 1991).

Behavior-coding studies by Oksenberg and others (1991), Loosveldt (1995), Dijkstra, Van der Veen and Van der Zouwen (1985) and Smit (1995) show that closed questions may easily lead to interviewer effects. Smit found that in 42 percent of the cases where interviewers had to read a list of response options, they presented fewer than all options. As suggested, this may be partly because respondents begin to answer before the interviewer is able to read the entire list of options. When a respondent answers the question before the interviewer presents all the options, the interviewer tends to read only a selection.

In chapter 5 we will see that closed questions as found in survey interviews are often formulated such that they implicitly cue respondents to take their turn before the interviewer reads the response options. The questionnaire tacitly guides both interviewer and respondent to speak at the same time.

As Fowler and Mangione (1986) point out, open questions also offer a potential for interviewer effects, caused by possible ambiguity concerning the kind of answer that will suffice. I will show that survey questions in general are scripted such that they create a response expectation that runs contrary to what the researchers had in mind.

If Fowler and Mangione are right, that open questions tend to be ambiguous with respect to the way respondents should deal with them, then field coded questions must be even more problematic. From the respondent's perspective, field coded questions are just open questions. For the interviewer, however, they are similar to closed questions. The significant difference is that field coded questions may generate answers that do not match the hidden response options. When receiving an unrecordable answer, the interviewer must probe for an answer that will match the pre-coded options. In Smit's study (1995) two field coded questions

6 Interaction and the standardized survey interview

led to a high degree of directive interviewer behavior, such that Smit advises against using field coded questions in survey research.

3.4 *Question wording and the interpretation of questions*

Belson (1981) convincingly shows the difficulty in phrasing questions that will be interpreted as intended by the researcher. He presented the following question to fifty-nine individuals: “How many days of the week do you usually watch television? I mean weekdays and Saturdays and Sundays, of course, and daytime viewing as well as evening viewing?” After respondents completed the questionnaire, a second interviewer tested the respondents’ understanding of the question, producing the following results:

- Fourteen out of fifty-nine failed to interpret this question as a request to mention a numerical answer.
- A large majority did not interpret “days of the week” as meant by the researchers.
- Five persons out of fifty-nine did not interpret “you” as referring to themselves only. They turned out to have treated “you” as “you and your wife,” “you and your family” or “your family.”
- Fifteen persons interpreted “watch television” as “have the TV set on,” rather than actually paying attention to a program.
- Twenty-eight out of fifty-nine did not interpret the term “usually” as intended.

Groves et al. (1992) asked respondents the closed-ended question, “Would you say your own health, in general, is excellent, good, fair, or poor?” It turned out that 43 percent of the men and 28 percent of the women rated their health condition as “excellent.” When respondents were asked via an open follow-up question whether they had compared their own health to that of others, it turned out that 12 percent said “yes,” and 88 percent said “no.” Asked whether they had compared their health now to it at an earlier age, 43 percent said “yes,” and 56 percent said “no.” These answers show that respondents evaluated their health condition from at least two different perspectives, which makes their answers incomparable.

Experimental studies show that small changes in question wording sometimes produce very large changes in response distribution. Butler and Kitzinger (1976) asked British people, “Do you accept the government’s recommendation that the United Kingdom should *come out of the Common Market*?” The difference between pro and con response was only 0.2 percent in favor of the Common Market position. However,

when the question read “Do you accept the government’s recommendation that the United Kingdom should *stay in* the Common Market?” the difference was 18.2 percent in favor of the pro position.

Rugg (1941) found a 21 percent difference in response distribution depending on whether the verb “allowing” or the verb “not forbidding” was used in a question on supporting freedom of speech. For other studies in the forbid/allow asymmetry, see Schuman and Presser (1981) on American data; Waterplas, Billiet, and Loosveldt (1988) on Belgian data; Hippler and Schwarz (1986) on German data; and Holleman (1996) on Dutch data.

The list of response options may contain a “don’t know” or “no-opinion” option. An experimental study by Schuman and Presser (1981) shows that 90 percent of the sample chose the “no opinion” option when it was offered to them. When this option was not offered, 69 percent volunteered they had no opinion on the issue. When interviewers vary in presenting this option, they create a bias.

Several studies show that vague quantifiers like “usually,” “many,” “some,” or “a few” are interpreted differently by different respondents (Bradburn and Miles 1979). Schaeffer (1991b) found a significant difference in the interpretation of phrases containing vague quantifiers for race, age, and level of education. The meaning of vague quantifiers differs also concerning the verbs and nouns that they quantify. The meaning of “many” in “having many children” and “having many books” may very well differ for one and the same respondent. Similarly, “watching television often” may well refer to more times than “going out often.” Similar problems exist for vague qualifiers, for example, “excellent,” “good,” “fair,” and “poor.”

Experimental evidence suggests that respondents often answer questions having little or no knowledge at all of the topic. In his review of the literature, Smith (1984) mentions the following findings: 96 percent of a sample offered opinions on the importance of a balanced federal budget. Only 3 percent of them had enough information on which to base an opinion. Ferber (1956) studied attitudinal questions and found that people expressed opinions they were not able to define. Converse (1970) would later refer to such opinions as non-attitudes. In fact, this notion may not quite cover the entire issue, as a Dutch field experiment by Van der Sloot and Spanjer (1998) makes clear. They presented respondents with a number of topically related opinion questions, some of which contained non-existent objects. The analysis of the transcripts makes clear that respondents may use the context to try to make sense of these terms (Schwartz 1996). In response to the question whether the (non-existent)

8 Interaction and the standardized survey interview

“televIEWer” was a practical telecom service, one respondent replied, “Is that a picture phone, or something?” Respondents may also think they misheard the interviewer. When the interviewer asked a question about the non-existent “NPM,” one person assumed the interviewer said, or meant to say, “NPB, the Netherlands Press Bureau.” This study also shows that respondents are willing to give an opinion about things they have no opinion about. Some of the respondents were very explicit about this. One person said, “I have no clue what it is, but just put no.”

3.5 *Standardized wording versus standardized meaning*

Fowler and Mangione (1990:136–7) state that questions should meet at least the following standards:

- Questions are fully scripted.
- Questions mean the same to every respondent.
- Answer options are read to all respondents.

The combination of the first two requirements suggests that standardized question wording and standardized meaning go together. However, when we look at the many studies on respondents interpreting survey questions, it will be clear that almost any word or concept used in questionnaires may be interpreted in more than one way. The main problem for survey methodology and for questionnaire designers is that these studies do not tell us how to phrase questions in an unambiguous way right from the start. This is because unambiguous questions hardly exist, due to the intrinsic ambiguity of language. The best that authors can come up with are pieces of advice on what questionnaire designers should avoid doing (e.g., Moser and Kalton 1978, Belson 1981).

The general advice to questionnaire designers is to pre-test their questions. However, when we look at, for example, Belson’s work, it is clear that a serious pre-test of a questionnaire is too difficult for most research because of time and money constraints. But suppose serious pre-testing were feasible, what could its practical use be? We might decide to explain one problematic question in more detail. A problem could then be that the question becomes far too long and complex. Molenaar (1982), who reviewed the relevant literature, found that longer questions increase the chance that response effects occur. Furthermore, explanation of the ambiguous question would contain ambiguous language too. The same problem appears when we decide to use filter questions.

The only sensible conclusion with regard to this issue, then, is that standardization of question wording does not necessarily imply standardization of meaning. The first person to point this out may be Lazarsfeld (1944), and since that time it has been stated again and again (Nuckols

1953; Cicourel 1964, 1982; Briggs 1986; Mishler 1986; Suchman and Jordan 1990a; Schaeffer 1991a; Foddy 1993). Remarkably, the ocean of studies in question interpretation has not led survey methodology to give up the idea of presenting respondents with standardized question wording. Standardization is, after all, the heart of survey methodology and market research, and for that reason it would not be easy to give it up.

In the next section I will discuss how survey methodology sees the role of the interviewer in the standardization of the measurement process. Since Fowler and Mangione are well-known protagonists of standardization, I will make extensive use of their 1990 study.

4 **Standardization of the interviewer–respondent interaction**

Because a change in the formulation or the order of questions and response options can have an effect on the responses, it is crucial that interviewers do not alter anything in the questionnaire. Interviewers, therefore, are to follow the rules of standardized interviewing. Fowler and Mangione (1990: 33) formulate the rules of standardized interviewing as follows:

- Read the questions exactly as worded.
- If the respondent's answer is incomplete or inadequate, probe for clarification or elaboration in a non-directive way.
- Record the answers without interviewer's discretion.
- Do not provide any positive or negative feedback regarding the specific content of respondent's answers.

Several studies in which interviewer–respondent interactions were coded show that interviewers often do not act according to these rules. For example, Brenner's (1982) analysis of tape-recorded interviews shows that more than 30 percent of all questions were not asked in the required manner. In a Dutch study, Dijkstra et al. (1985) found that interviewers altered the essential content or the meaning of the written questions to the extent that 40 percent of all answers are not to be trusted. For similar findings, see Belson (1965), Bradburn and Sudman (1979), Cannell, Fowler, and Marquis (1968), Fowler and Mangione (1986), Cannell and Oksenberg (1988).

Why are interviewers often not rule-obedient? Fowler and Mangione (1990: 33–4) summarize the obstacles to carrying out an interview in a standardized way. One is that interviewers like to be considerate and responsive to respondents, and feel this to be in conflict with standardized behavior.

Another possible problem is that the questionnaire may be inadequate.

For example, a question may be ambiguous or difficult to read aloud. Such potential problems can, to some degree, be solved by pre-testing the questionnaire (see Czaja and Blair 1996 on pre-testing procedures). Furthermore, respondents may not understand what they are expected to do. They often do not act their roles as prescribed. Fowler and Mangione advocate that interviewers explain clearly to respondents the rules of the game, and they present the following text to be read:

Since many people have never been in an interview exactly like this, let me read you a paragraph that tells a little bit about how it works. I am going to read you a set of questions exactly as they are worded so that every respondent in the survey is answering the same questions. You'll be asked to answer two kinds of questions. In some cases, you'll be asked to answer in your own words. For those questions, I will have to write down your answers word for word. In other cases, you will be given a list of answers and asked to choose the one that fits best. If at any time during the interview you are not clear about what is wanted, be sure to ask me. (1990: 51)

Cannell et al. (1981) found that reporting improves when the respondent's role is made clear by explicit instruction. In this book it will become clear that such an introduction will hardly solve the problem of respondents failing to do what is expected from them.

5 Probing inadequate answers

When respondents do not provide satisfactory answers, interviewers must probe for more adequate or more complete ones. And they should do this in a non-directive way, that is, interviewers' probes should not increase the chance of one answer over another. As Fowler and Mangione (1990) say, each probe that can be answered by "yes" or "no" is directive. They propose a number of rules that interviewers should follow when probing answers:

- When respondents miss part of the question, interviewers should reread the entire question.
- When respondents do not understand a term or concept in the question, for example, because a term turns out to be ambiguous, interviewers are not allowed to provide an explanation. They can reread the definition, if the questionnaire provides one. Otherwise respondents must be told to interpret the question to their own liking and answer the question.
- When probing answers, interviewers should make sure that answers meet the questions' objectives. It is the interviewers' task to make sure that respondents indeed provide a numerical answer, or choose one of the response options read to them.