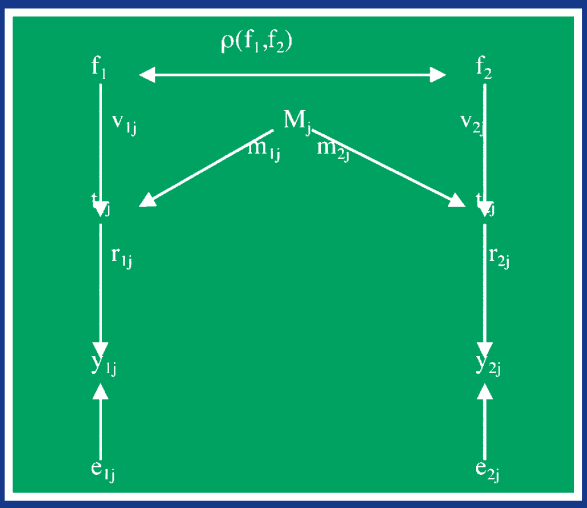
Wiley Series in Survey Methodology

Design, Evaluation, and Analysis of Questionnaires for Survey Research

SECOND EDITION



Willem E. Saris
Irmtraud N. Gallhofer

WILEY

DESIGN, EVALUATION, AND ANALYSIS OF QUESTIONNAIRES FOR SURVEY RESEARCH

WILEY SERIES IN SURVEY METHODOLOGY
Established in Part by WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors: *Mick P. Couper, Graham Kalton, J. N. K. Rao, Norbert Schwarz, Christopher Skinner*
Editor Emeritus: *Robert M. Groves*

A complete list of the titles in this series appears at the end of this volume.

DESIGN, EVALUATION, AND ANALYSIS OF QUESTIONNAIRES FOR SURVEY RESEARCH

Second Edition

WILLEM E. SARIS AND IRMTRAUD N. GALLHOFER
Research and Expertise Centre for Survey Methodology
Universitat Pompeu Fabra
Barcelona, Spain

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Saris, Willem E.
Design, evaluation, and analysis of questionnaires for survey research / Willem E. Saris, Irmtraud Gallhofer. – Second Edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-63461-5 (cloth)
1. Social surveys. 2. Social surveys–Methodology. 3. Questionnaires. 4. Interviewing. I. Title.
HN29.S29 2014
300.72′3–dc23

2013042094

CONTENTS

PREFACE TO THE SECOND EDITION

The most innovative contribution of the first edition of the book was the introduction of a computer program (SQP) for predicting the quality of survey questions, created on the basis of analyses of 87 multitrait–multimethod (MTMM) experiments. At that time (2007), this analysis was based on 1067 questions formulated in three different languages: English, German, and Dutch. The predictions were therefore also limited to questions in these three languages.

The most important rationale for this new edition of the book is the existence of a new SQP 2.0 program that provides predictions of the quality of questions in more than 22 countries based on a database of more than 3000 extra questions that were evaluated in MTMM experiments to determine the quality of the questions. The new data was collected within the European Social Survey (ESS). This research has been carried out since 2002 every two years in 36 countries. In each round, four to six experiments were undertaken to estimate the quality of approximately 50 questions in all countries and in their respective languages. This means that the new program has far more possibilities to predict the quality of questions in different languages than its predecessor, which was introduced in the first edition of the book.

Another very important reason for a new edition of the book is also related to the new program. Whereas the earlier version had to be downloaded and used on the same PC, the new one is an Internet program with a connected database of survey questions. These contain all questions used in the old experiments as well as the new experiments, but equally, all questions asked to date in the ESS. This means that the SQP database contains more than 60,000 questions in all languages used in the ESS and elsewhere. The number of questions will grow in three ways: (1) by way of the new studies done by the ESS, which adds another 280 questions phrased in all of its working languages used in each round; (2) as a result of the new studies added to the

database by other large-scale cross-national surveys; and (3) thanks to the introduction of new questions by researchers who use the program in order to evaluate the quality of their questions. In this way, the SQP program is *a continuously growing database of survey questions in most European languages with information about the quality of the questions and about the possibility for evaluating the quality of questions that have not yet been evaluated*. The program will thus be a permanently growing source of information about survey questions and their quality. To our knowledge, there is no other program that exists to date that offers the same possibilities.

We have used this opportunity to improve two chapters based on the comments we have received from program users. This is especially true for Chapter 1 and Chapter 15. Furthermore, we decided to adjust Chapters 12 and 16 on the basis of new developments in the field.

Willem E. Saris
Irmtraud Gallhofer

PREFACE

Designing a survey involves many more decisions than most researchers realize. Survey specialists, therefore, speak of the art of designing survey questions (Payne 1951). However, this book introduces the methods and procedures that can make questionnaire design a scientific activity. This requires knowledge of the consequences of the many decisions that researchers take in survey design and how these decisions affect the quality of the questions.

It is desirable to be able to evaluate the quality of the candidate questions of the questionnaire before collecting the data. However, it is very tedious to manually evaluate each question separately on all characteristics mentioned in the scientific literature that predicts the quality of the questions. It may even be said that it is impossible to evaluate the effect of the combination of all of these characteristics. This would require special tools that did not exist so far. A computer program capable of evaluating all the questions in a questionnaire according to a number of characteristics and providing an estimate of the quality of the questions based on the coded question characteristics would be very helpful. This program could be a tool for the survey designer in determining, on the basis of the computer output, which questions in the survey require further study in order to improve the quality of the data collected.

Furthermore, after a survey is completed, it is useful to have information about the quality of the data collected in order to correct for errors in the data. Therefore, there is a need for a computer program that can evaluate all questions of a questionnaire based on a number of characteristics and provide an estimate of the quality of the questions. Such information can be used to improve the quality of the data analysis.

In order to further such an approach, we have

1. Developed a system for coding characteristics of survey questions and the more general survey procedure;
2. Assembled a large set of studies that used multitrait–multimethod (MTMM) experiments to estimate the reliability and validity of questions;
3. Carried out a meta-analysis that relates these question characteristics to the reliability and validity estimates of the questions;
4. Developed a semiautomatic program that predicts the validity and reliability of new questions based on the information available from the meta-analysis of MTMM experiments.

We think that these four steps are necessary to change the development of questionnaires from an "art" into a scientific activity.

While this approach helps to optimize the formulation of a single question, it does not necessarily improve the quality of survey measures. Often, researchers use complex concepts in research that cannot be measured by a single question. Several indicators are therefore used. Moving from complex concepts to a set of questions that together may provide a good measure for the concept is called operationalization. In order to develop a scientific approach for questionnaire design, we have also provided suggestions for the *operationalization* of complex concepts.

The purpose of the book is, first, to specify a three-step procedure that will generate questions to measure the complex concept defined by the researcher. The approach of operationalization is discussed in Part I of the book.

The second purpose of the book is to introduce to survey researchers the different choices they can make and are making while designing survey questionnaires, which is covered in Part II of the book.

Part III discusses quality criteria for survey questions, the way these criteria have been evaluated in experimental research, and the results of a meta-analysis over many of such experiments that allow researchers to determine the size of the effects of the different decisions on the quality of the questions.

Part IV indicates how all this information can be used efficiently in the design and analysis of surveys. Therefore, the first chapter introduces a program called "survey quality predictor" (SQP), which can be used for the prediction of the quality of survey items on the basis of cumulative information concerning the effect of different characteristics of the different components of survey items on the data quality. The discussion of the program will be specific enough so that the reader can use it to improve his/her own questionnaires.

The information about data quality can and should also be used after a survey has been completed. Measurement error is unavoidable, and this information is useful for how to correct it. The exact mechanics of it are illustrated in several chapters of Part IV. We start out by demonstrating how this information can be applied to estimate the quality of measures of complex concepts, followed by a discussion on how to correct for measurement error in survey research. In the last chapter, we discuss how one can cope with measurement error in cross-cultural research.

In general, we hope to contribute to the scientific approach of questionnaire design and the overall improvement of survey research with the book.

ACKNOWLEDGMENTS

This second edition of the book would not have been possible without the dedicated cooperation in the data collection by the national coordinators of the ESS in the different countries and the careful work of our colleagues in the central coordinating team of the ESS.

All the collected data has been analyzed by a team of dedicated researchers of the Research and Expertise Centre for Survey Methodology, especially Daniel Oberski, Melanie Revilla, Diana Zavala Rojas, and our visiting scholar Laur Lilleoja. We can only hope that they will continue their careful work in order to improve the predictions of SQP even more in the future. The program would not have been created without the work of two programmers Daniel Oberski and Tom Grüner.

Finally, we would like to thank our publisher Wiley for giving us the opportunity to realize the second edition of the book. A very important role was also played by Maricia Fischer-Souan who was able to transform some of our awkward English phrases into proper ones.

Last but not least, we would like to thank the many scholars who have commented on the different versions of the book and the program. Without their stimulating support and criticism, the book would not have been written.

INTRODUCTION

In order to emphasize the importance of survey research for the social, economic, and behavioral fields, we have elaborated on a study done by Stanley Presser, originally published in 1984. In this study, Presser performed an analysis of papers published in the most prestigious journals within the scientific disciplines of economics, sociology, political science, social psychology, and public opinion (or communication) research. His aim was to investigate to what extent these papers were based on data collected in surveys.

Presser did his study by coding the data collection procedures used in the papers that appeared in the following journals. For the economics field, he used the *American Economic Review*, the *Journal of Political Economy*, and the *Review of Economics and Statistics*. To represent the sociology field, he used the *American Sociological Review*, the *American Journal of Sociology*, and *Social Forces* and, for the political sciences, the *American Journal of Political Science*, the *American Political Science Review*, and the *Journal of Politics*. For the field of social psychology, he chose the *Journal of Personality and Social Psychology* (a journal that alone contains as many papers as each of the other sciences taken together). Finally, for public opinion research, the *Public Opinion Quarterly* was elected. For each selected journal, all papers published in the years 1949–1950, 1964–1965, and 1979–1980 were analyzed.

We have updated Presser's analysis of the same journals for the period of 1994–1995, a period that is consistent with the interval of 15 years to the preceding measurement. Presser (1984: 95) suggested using the following definition of a survey:

Design, Evaluation, and Analysis of Questionnaires for Survey Research, Second Edition.
Willem E. Saris and Irmtraud N. Gallhofer.

TABLE I.1 Percentage of articles using survey data by discipline and year (number of articles excluding data from statistical offices in parentheses)

	Period			
Discipline	1949–1950	1964–1965	1979–1980	1994–1995
Economics	5.7% (141)	32.9% (155)	28.7% (317)	(20.0%) 42.3% (461)
Sociology	24.1% (282)	54.8% (259)	55.8% (285)	(47.4%) 69.7% (287)
Political science	2.6% (114)	19.4% (160)	35.4% (203)	(27.4%) 41.9% (303)
Social psychology	22.0% (59)	14.6% (233)	21.0% (377)	(49.0%) 49.9% (347)
Public opinion	43.0% (86)	55.7% (61)	90.6% (53)	(90.3%) 90.3% (46)

> …any data collection operation that gathers information from human respondents by means of a standardized questionnaire in which the interest is in aggregates rather than particular individuals. (…) Operations conducted as an integral part of laboratory experiments are not included as surveys, since it seems useful to distinguish between the two methodologies. The definition is silent, however, about the method of respondent selection and the mode of data collection. Thus, convenience samples as well as census, self-administered questionnaires as well as face-to-face interviews, may count as surveys.

The results obtained by Presser, and completed by us for the years 1994–1995, are presented in Table I.1. For completing the data, we stayed consistent with the procedure used by Presser except in one point: we did not automatically subsume studies performed by organizations for official statistics (statistical bureaus) under the category "surveys." Our reason was that at least part of the data collected by statistical bureaus is based on administrative records and not collected by survey research as defined by Presser. Therefore, it is difficult to decide on the basis of the description of the data in the papers whether surveys have been used. For this reason, we have not automatically placed this set of papers, based on studies by statistical bureaus, in the class of survey research.

The difference in treating studies from statistical bureaus is reflected in the last column of Table I.1, relating to the years 1994–1995. We first present (within parentheses) the percentage of studies using survey methods based on samples (our own classification). Next, we present the percentages that would be obtained if all studies conducted by statistical bureaus were automatically subsumed under the category survey (Presser's approach).

Depending on how the studies of the statistical offices are coded, the proportion of survey research has increased, or slightly decreased, over the years in economics, sociology, and political science. Not surprisingly, the use of surveys in public opinion research is still very high and stable.

TABLE I.2 Use of different data collection methods in different disciplines as found in the major journals in 1994–1995 expressed in percentages with respect to the total number of empirical studies published in these years

	Disciplines				
Method	Economics	Sociology	Political science	Psychology	Public opinion
Survey	39.4	59.6	28.9	48.7	95.0
Experimental	6.0	1.7	5.4	45.6	5.0
Observational	3.2	0.6	31.9	4.1	0.0
Text analysis	6.0	4.6	7.2	0.6	0.0
Statistical data	45.4	33.5	26.6	9.0	0.0

Most remarkable is the increase of survey research in social psychology: the proportion of papers using survey data has more than doubled over the last 15-year interval. Surprisingly, this outcome contradicts Presser's assumption that the limit of the survey research growth in the field of social psychology might already have been reached by the end of the 1970s, due to the "field's embracing the laboratory/experimental methodology as the true path to knowledge."

Presser did not refer to any other method used in the papers he investigated, except for the experimental research of psychologists. For the papers published in 1994–1995, we, however, also categorized nonsurvey methods of the papers. Moreover, we checked whether any empirical data were employed in the same papers.

In economics, sociology, and political science, many papers are published that are purely theoretical, that is, formulating verbal or mathematical theories or discussing methods. In economics, this holds for 36% of the papers; in sociology, this figure is 26%; and in political science, it is 34%. In the journals representing the other disciplines, such papers have not been found for the period analyzed.

Given the large number of theoretical papers, it makes sense to correct the percentages of Table I.1 by ignoring the purely theoretical papers and considering only empirical studies. The results of this correction for 1994–1995 are presented in Table I.2.

Table I.2 shows the overwhelming importance of the survey research methodology for public opinion research but also for sociology and even for social psychology. For social psychology, the survey method is at least as important as the experimental design, while hardly any other method is employed. In economics and sociology, existing statistical data also are frequently used, but it has to be considered that these data sets themselves are often collected through survey methods.

The situation in political science in the period of 1994–1995 is somewhat different, although political scientists also use quite a number of surveys and statistical data sets based on surveys; they also make observations in many papers of the voting behavior of representatives.

We can conclude that survey research has become even more important than it was 15 years ago, as shown by Presser. All other data collection methods are only used infrequently with the exception of what we have called "statistical data." These data

are collected by statistical bureaus and are at least partially based on survey research and on administrative records. Observations, in turn, are used especially in the political sciences for researching voting behavior of different representative bodies, but hardly in any other science. The psychologists naturally use experiments but with less frequency than was expected from previous data. In communication science, experiments are also utilized on a small scale. All in all, this study clearly demonstrates the importance of survey research for the fields of the social and behavioral sciences.

I.1 DESIGNING A SURVEY

As a survey is a rather complex procedure to obtain data for research, in this section, we will briefly discuss a number of decisions a researcher has to take in order to design a survey.

I.1.1 Choice of a Topic

The first choice to be made concerns the substantive research in question. There are many possibilities, depending on the state of the research in a given field what kind of research problem will be identified. Basic choices are whether one would like to do a *descriptive* or *explanatory* study and in the latter case whether one would like to do *experimental* research or *nonexperimental* research.

Survey research is often used for descriptive research. For example, in newspapers and also in scientific journals like *Public Opinion Quarterly*, many studies can be found that merely give the distribution of responses of people on some specific questions such as satisfaction with the economy, government, and functioning of the democracy. Many polls are done to determine the popularity of politicians, to name just a few examples.

On the other hand, studies can also be done to determine the reasons for the satisfaction with the government or the popularity of a politician. Such research is called *explanatory research*. The class of explanatory studies includes nonexperimental as well as experimental studies in a laboratory. Normally, we classify research as *survey research* if large groups of a population are asked questions about a topic. Therefore, even though laboratory experiments employ questionnaires, they are not treated as surveys in this book. However, nowadays experimental research can also be done with survey research. In particular, computer-assisted data collection facilitates this kind of research by random assignment procedures (De Pijper and Saris 1986; Piazza and Sniderman 1991), and such research is included here as survey research. The difference between the two experimental designs is where the emphasis is placed, either on the data of individuals or small groups or on the data of some specified population.

I.1.2 Choice of the Most Important Variables

The second choice is that of the variables to be measured. In the case of a descriptive study, the choice is rather simple. It is directly determined by the purpose of the study. For example, if a study is measuring the satisfaction of the population

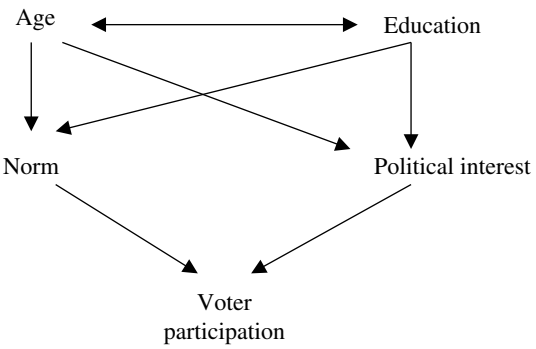


FIGURE I.1 A model for the explanation of participation in elections by voting.

with the government, it is clear that questions should be asked about the "satisfaction with the government."

On the other hand, to study what the effects of different variables are on participation in elections, the choice is not so clear. In this case, it makes sense to develop an inventory of possible causes and to develop from that list a preliminary model that indicates the relationships between the variables of interest. An example is given in Figure I.1. We suppose that two variables have a *direct effect* on "participation in elections" (voter participation): "political interest" and "the adherence to the norm that one should vote."

Furthermore, we hypothesize that "age" and "education" have a direct influence on these two variables but only an *indirect effect* on "participation in elections." One may wonder why the variables age and education are necessary in such a study if they have no direct effect on "voter participation." The reason is that these variables cause a relationship between the "norm" and "voter participation" and, in turn, between "political interest" and "voter participation." Therefore, if we use the correlation between, for example, "political interest" and "voter participation" as the estimate of the effect of "political interest," we would overestimate the size of the effect because part of this relationship is a "spurious correlation" due to "age" and "education."

For more details on this issue, we recommend the following books on causal modeling by Blalock (1964), Duncan (1975), and Saris and Stronkhorst (1984). Therefore, in this research, one not only has to introduce the variables "voter participation," "political interest," and "adherence to the norm" but also "age" and "education" as well as all other variables that generate spurious correlation between the variables of interest.

I.1.3 Choice of a Data Collection Method

The third choice to be made concerns the data collection method. This is an important choice related to costs, question formulation, and quality of data. Several years ago, the only choices available were between personal interviews (face-to-face interviews),

telephone interviews, and mail surveys, all using paper questionnaires. A major difference in these methods was the presence of the interviewer in the data collection process. In personal interviews, the interviewer is physically present; in telephone interviewing, the interviewer is at a distance and the contact is by phone; while in mail surveys, the interviewer is not present at all. Nowadays, each of these modes of data collection can also be computerized by computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), and computer-assisted self-interviewing (CASI) or Web surveys.

As was mentioned, these modes of data collection differ in their cost of data collection, where personal interviewing is the most expensive, telephone interviewing is less expensive, and mail interviewing is the cheapest. This holds true even with the aid of the computer. The same ordering can be specified for the response that one can expect from the respondents although different procedures have been developed to reduce the nonresponse (Dillman 2000).

Besides the aforementioned differences, there is a significant amount of literature on the variances in data quality obtained from these distinct modes of data collection. We will come back to this issue later in the book, but what should be clear is that the different modes require a corresponding formulation of the questions, and due to these differences in formulation, differences in responses can also be expected. Therefore, the choice of the mode of data collection is of critical importance not only for the resulting data quality but also for the formulation of the questions, which is the fourth decision to be made while designing a survey.

I.1.4 Choice of Operationalization

Operationalization is the translation of the concepts to the questions. Most people who are not familiar with designing questionnaires think that making questionnaires is very simple. This is a common and serious error. To demonstrate our point, let us look at some very simple examples of questions:

I.1 Do you like football?

Most women probably answered the question: *Do you like to watch football on TV?*

Most young men will answer the question: *Do you like to play football?*

Some older men will answer the former question, some others the latter one, depending on whether they are still playing football.

This example shows that the interpretation of the question changes for the age and gender of the respondents.

Let us look at another example of a question that was frequently asked in 2003:

I.2a Was the invasion of Iraq in 2003 a success?

In general, the answer to this question is probably "yes." President Bush declared the war over in a relatively short time. But the reaction would have been quite different in 2004 if it had been asked:

I.2b Is the invasion of Iraq in 2003 a success?

Probably, the answer would be "no" for most people because after the end of the war, the initial problem was not solved.

While there is only a one word difference in these questions, the responses of the people would have been fundamentally different because in the first question (I.2a), people answer a question about the invasion, but in the second question (I.2b), they shift the object to evaluating the consequences of the invasion at that later point in time.

Given that such simple questions can already create a problem, survey specialists speak of "the art of asking questions" (Payne 1951; Dillman 2000: 78). We think that there is a third position on this issue: that it is possible to develop scientific methods for questionnaire design. In designing a question, many decisions are made. If we know the consequences of these decisions on the quality of the responses, then we can design *optimal questions* using a scientific method.

Now, let us consider some decisions that have to be made while designing a question.

Decision 1: *Subject and Dimension*
A researcher has to choose a subject and a dimension on which to evaluate the subject of the question. Let us expand on examples I.2a and I.2b:

I.2c *Was the invasion a success?*
I.2d *Was the invasion justified?*
I.2e *Was the invasion important?*

For examples I.2c–I.2e, there are many more choices possible, but what is done here is that the *subject* (the invasion) has been kept the same and the *dimension* on which people have to express their answer (*concept* asked) changes. The researcher has to make the choice of the dimension or concept depending on the purpose of the study.

Decision 2: *Formulation of the Question*
Many different formulations of the same question are also possible. For example:

I.2f *Was the invasion a success?*
I.2g *Please tell me if the invasion was a success.*
I.2h *Now, I would like to ask you whether the invasion was a success.*
I.2i *Do you agree or not with the statement: the invasion was a success.*

Again, there are many more formulation choices possible, as we will show later.

Decision 3: *The Response Categories*
The next decision is choosing an appropriate response scale. Here, again are some examples:

I.2j	*Was the invasion a success?*	*Yes/no*
I.2k	*How successful was the invasion?*	*Very much/quite/a bit/not at all*
I.2 l	*How successful was the invasion?*	*Express your opinion with a number between 0 and 100 where 0 = no success at all and 100 = complete success*

Again, there are many more formulation options, as we will discuss later in the book.

Decision 4: *Additional Text*
Besides the question and answer categories, it is also possible to add:

- An introduction
- Extra information
- Definitions
- Instructions
- A motivation to answer

It is clear that the formulation of a single question has many possibilities. The study of these decisions and their consequences on the quality of the responses will be the main topic of this book. But before we discuss this issue, we will continue with the decisions that have to be made while designing a survey study.

I.1.5 Test of the Quality of the Questionnaire

The next step in designing a survey study is to conduct a check of the quality of the questionnaire. Some relevant checks are:

- Check on face validity
- Control of the routing in the questionnaire
- Prediction of quality of the questions with some instrument
- Use of a pilot study to test the questionnaire

It is always necessary to ask yourself and other people whether the concepts you want to measure are really measured by the way the questions are formulated. It is also necessary to control for the correctness of all routings in the questionnaire. This is especially important in computer-assisted data collection because otherwise the respondent or interviewer can be guided completely in the wrong direction, which normally leads to incomplete responses.

There are also several approaches developed to control the quality of questions. This can be done by an expert panel (Presser and Blair 1994) or on the basis of a coding scheme (Forsyth et al. 1992; Van der Zouwen 2000) or by using a computer program (Graesser et al. 2000a, b). Another approach that is now rather popular is to present respondents with different formulations of a survey item in a laboratory setting in order to understand the effect of wording changes (Esposito et al. 1991; Esposito and Rothgeb 1997). For an overview of the different possible cognitive approaches to the evaluation of questions, we recommend Sudman et al. (1996).

In this book, we will provide our own tool, namely, survey quality predictor (SQP), which can be used to predict the quality of questions before they are used in practice.

I.1.6 Formulation of the Final Questionnaire

After corrections in the questionnaire have been made, the ideal scenario would be to test the new version again. With respect to the routing of computer-assisted data collection, that is certainly the case because of the serious consequences if something is off route. Another is to ensure that people actually understand a question better after correction. However, it will be clear that there is a limit to the iteration of tests and improvements.

Another issue is that the final layout of the questionnaire has to be decided on. This holds equally for both the paper-and-pencil approach and for questionnaires designed for computer-assisted data collection. However, research has only started on the effects of the layout on quality of the responses. For further analysis of the issue, see Dillman (2000).

After all these activities, the questionnaires can be printed if necessary to follow through with the data collection.

So far, we have concentrated on the design of the questionnaire. There is, however, another line of work that also has to be done. This concerns the selection of a population and sampling design and organization of the fieldwork, which will be discussed in the subsequent sections.

I.1.7 Choice of Population and Sample Design

With all survey research, a decision about what *population* to report on has to be made. One possible issue to consider is whether to report about the population of the country as a whole or about a specific subgroup. This decision is important because without it a sampling design cannot be specified. *Sampling* is a procedure to select a limited number of units from a population in order to describe this population. From this definition, it is clear that a population has to be selected first.

The sampling should be done in such a way that the researcher has no influence on the selection of the respondents; otherwise, the researcher can influence the results. The recommended procedure to satisfy this requirement is to select the respondents at random. Such samples based on a selection at random are called *random samples*.

If a random sampling procedure is used with a known selection probability for all respondents (not zero and not necessarily equal for all people), then it is in principle possible to *generalize* from the sample results to the population. The precision of the statements one can make about the population depends on the design of the sample and the size of the sample.

In order to draw a sample from a population, a *sampling frame* such as a list of names and addresses of potential respondents is needed. This can be a problem for specific populations, but if such a list is missing, there are also procedures to create a sampling frame. For further details, we refer to the standard literature in the area (Kish 1965; Cochran 1977; Kalton 1983). It should, however, be clear that this is a very important part of the design of the survey instrument that has to be worked out very carefully and on the basis of sufficient knowledge of the topic.

I.1.8 Decide about the Fieldwork

At least as important as the design of the sample is the design of the fieldwork. This stage determines the amount of cooperation and refusals from respondents and the quality of the work of the interviewers. In order to generate an idea of the complexity of this task, we provide an overview of the decisions that have to be made:

- Number of interviews for each interviewer
- Number of interviewers
- Recruitment of interviewers: where, when, and how
- How much to pay: per hour/per interview
- Instruction: kind of contacts, number of contacts, when to stop, and administration
- Control procedures: interviews done/not done
- Registration of incoming forms
- Coding of forms
- Necessary staff

All these decisions are rather complex and require special attention in survey research, which are beyond the scope of this book.

I.1.9 What We Know about These Decisions

In his paper mentioned at the beginning of this introduction, Presser (1984) complained that, in contrast with the importance of the survey method, methodological research was directed mainly at statistical analysis and not at the methods of data collection itself. That his observation still holds can be seen if one looks at the high proportion of statistical papers published in *Sociological Methodology* and in *Political Analysis*, the two most prestigious methodological outlets in the social sciences. However, we think that the situation has improved over the last 15 years in that research has been done directed at the quality of the survey method. The following section will be a brief review of this research.

In psychology, large sets of questions are used to measure a concept. The quality of these so-called tests are normally evaluated using factor analysis, classical test theory models, and reliability measures like Cronbach's α or item response theory (IRT) models. In survey research, such large sets of questions are not commonly used. Heise (1969) presented his position for a different approach. He argued that the questions used by sociologists and political scientists cannot be seen as alternative measures for the same concept as in psychology. Each question measures a different concept, and therefore, a different approach for the evaluation of data quality is needed. He suggested the use of the quasi-simplex models, evaluating the quality of a single question in a design using panel studies. Saris (1981) showed that different questions commonly used for the measurement of "job satisfaction" cannot be seen as indicators of the same concept. Independently of these theoretical arguments, survey researchers are frequently using single questions as indicators for the concepts they want to measure.