2004

new technologies in data collection, questionnaire design and quality

datuak jasotzeko, galdesortak diseinatzeko eta inkesten kalitaterako metodologia berriak

nuevas metodologías en la recogida de datos, diseño de cuestionarios y calidad en las encuestas

EDITH D. DE LEEUW

44

nazioarteko estatistika mintegia euskadin

seminario internacional de estadística en euskadi

2004

new technologies in data collection, questionnaire design and quality

datuak jasotzeko, galdesortak diseinatzeko eta inkesten kalitaterako metodologia berriak

nuevas metodologías en la recogida de datos, diseño de cuestionarios y calidad en las encuestas

EDITH D. DE LEEUW

44

Eustat

***Lanketa** / Elaboración:*

Euskal Estatistika Erakundea
Instituto Vasco de Estadística (EUSTAT)

***Argitalpena** / Edición:*

Euskal Estatistika Erakundea
Instituto Vasco de Estadística
Donostia-San Sebastián, 1 - 01010 Vitoria-Gasteiz

***Ale-kopurua** / Tirada:*
500 **ale** / ejemplares

V-2005

Fotokonposaketarako tratamendu informatikoa:
Tratamiento informático de fotocomposición:
Estudios Gráficos ZURE S.A.
Ctra. Lutxana-Asua, 24 A
Erandio-Goikoa (BIZKAIA)

Inprimaketa eta Koadernaketa:
Impresión y Encuadernación:
Estudios Gráficos ZURE S.A.
Ctra. Lutxana-Asua, 24 A
Erandio-Goikoa (BIZKAIA)

I.S.B.N.: 84-7749-417-7

AURKEZPENA

Nazioarteko Estatistika Mintegia antolatzean, hainbat helburu bete nahi ditu EUSTAT-Euskal Estatistika Erakundeak:

- Unibertsitatearekiko eta, batez ere, Estatistika-Sailekiko lankidetza bultzatzea.
- Funtzionarioen, irakasleen, ikasleen eta estatistikaren alorrean interesatuta egon daitezkeen guztien lanbide-hobekuntza erraztea.
- Estatistika alorrean mundu mailan abangoardian dauden irakasle eta ikertzaile ospetsuak Euskadira ekartzea, horrek eragin ona izango baitu, zuzeneko harremanei eta esperientziak ezagutzeari dagokienez.

Jarduera osagarri gisa, eta interesatuta egon litezkeen ahalik eta pertsona eta erakunde gehienetara iristearren, ikastaro horietako txostenak argitaratzea erabaki dugu, beti ere txostengilearen jatorrizko hizkuntza errespetatuz; horrela, gai horri buruzko ezagutza gure herrian zabaltzen laguntzeko.

Vitoria-Gasteiz, 2005eko Maiatza

JOSU IRADI ARRIETA
EUSTATeko Zuzendari Nagusia

PRESENTATION

In promoting the International Statistical Seminars, EUSTAT-The Basque Statistics Institute wishes to achieve several aims:

- Encourage the collaboration with the universities, especially with their statistical departments.
- Facilitate the professional recycling of civil servants, university teachers, students and whoever else may be interested in the statistical field.
- Bring to the Basque Country illustrious professors and investigators in the vanguard of statistical subjects, on a worldwide level, with the subsequent positive effect of encouraging direct relationships and sharing knowledge of experiences.

As a complementary activity and in order to reach as many interested people and institutions as possible, it has been decided to publish the papers of these courses, always respecting the original language of the author, to contribute in this way towards the growth of knowledge concerning this subject in our country.

Vitoria-Gasteiz, May 2005

JOSU IRADI ARRIETA
General Director of EUSTAT

PRESENTACIÓN

Al promover los Seminarios Internacionales de Estadística, el EUSTAT-Instituto Vasco de Estadística pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a ilustres profesores e investigadores de vanguardia en materia estadística, a nivel mundial, con el consiguiente efecto positivo en cuanto a la relación directa y conocimiento de experiencias.

Como actuación complementaria y para llegar al mayor número posible de personas e Instituciones interesadas, se ha decidido publicar las ponencias de estos cursos, respetando en todo caso la lengua original del ponente, para contribuir así a acrecentar el conocimiento sobre esta materia en nuestro País.

Vitoria-Gasteiz, Mayo 2005

JOSU IRADI ARRIETA
Director General de EUSTAT

BIOGRAFI OHARRAK

EDITH D. DE LEEUW doktorea Herbehereetako Utrechteko Unibertsitateko Metodologia eta Estatistika saileko irakasle elkartua eta metodologia eta estatistikan aholkulari diren MethodikA Amsterdam-eko zuzendaria da. Washingtoneko Estatuko Unibertsitateko Ekonomia- eta Gizarte-zientzietako Ikerketa Zentroan Fulbright bekaduna Dillman irakaslearekin, eta Gizarte-estatistiketako Programako bisitari Los Angeleseko Californiako Unibertsitatean. Psikometria eta Soziometria Erakundeko kidea (IOPS).

De Leeuw doktorea Journal of Official Statistics-en (JOS) editore elkartua da, eta aurretik, International Sociological Association (ISA) elkarteko Logika eta Metodologia batzordeko idazkari betearazlea eta presidentea, eta Netherlands Society for Statistics and Operations Research-eko Gizarte-zientzietako sailekoa ere bai. Inkesten eta estatistikaren metodologiaren inguruko hitzaldiak antolatu ditu, nazionalak zein nazioartekoak. Utrechteko International University College eta ISR summer school-en eskolak ematen ditu. Estatistika Ofizialerako TES ikastaroetan irakasle izan da.

BIOGRAPHICAL SKETCH

EDITH D. DE LEEUW, Ph.D., is associate professor of survey methodology at the Department of Methodology and Statistics at Utrecht University, the Netherlands, and director of MethodikA Amsterdam, consultants in methodology and statistics. She was a Fulbright scholar with professor Dillman at the Social and Economic Sciences Research Center of Washington State University and visiting scholar of the Program on Social Statistics at UCLA. She is a fellow of the Joint Universities Institute for Psychometrics and Sociometrics (IOPS).

Dr de Leeuw is associate editor of the Journal of Official Statistics and was executive secretary and president of the international research committee on logic and methodology of the International Sociological Association (ISA), and of the Section on Social Science of the Netherlands Society for Statistics and Operations Research. She has organized both national and international conferences on survey methodology and statistics. She teaches at the International University College Utrecht and at the ISR summer school and has been a teacher at TES-courses for Official Statistics..

NOTAS BIOGRÁFICAS

EDITH D. DE LEEUW, Ph.D., es profesora asociada de metodología de encuestas en el Departamento de Metodología y Estadística de la Universidad de Utrech, Países Bajos, y directora de MethodikA Amsterdam, asesores en metodología y estadística. Beca Fulbright con el profesor Dillman en el Centro de Investigación de Ciencias Económicas y Sociales de la Universidad Estatal de Washington y visitante del Programa en Estadísticas Sociales en la Universidad de California en Los Ángeles. Miembro del Instituto de Psicometría y Sociometría (IOPS).

La doctora de Leeuw es editora asociada del Journal of Official Statistics (JOS), y antes secretaria ejecutiva y presidenta del comité de lógica y metodología de la International Sociological Association (ISA), y de la sección de Ciencias Sociales de la Netherlands Society for Statistics and Operations Research. Ha organizado conferencias nacionales e internacionales sobre metodología de encuestas y estadística. Es profesora en el International University College Utrecht y en el ISR summer school. Ha sido profesora de cursos TES para Estadística Oficial.

CONTENTS

1. Introduction

In December 2004, I taught a EUSTAT seminar on "New Technologies in Data Collection, Questionnaire Design, and Quality" in Bilbao, Spain. The course was designed for experienced survey researchers, as well as for those with little practical experience in survey methods. The purpose of my seminar and teaching philosophy was to provide participants with a better understanding of survey data quality and provide them with an efficient and up-to-date tool chest to design and implement high quality surveys. During the course, plenary lectures were alternated with short exercises and question-and answer sessions. Or in other words, tools were presented and participants learned to use these tools.

The seminar started with an introduction of total survey error and data quality. An overview of data collection methods was given and the influence of data collection and technology on quality was discussed. In the second part of the seminar, special attention was given to questionnaire design and optimizing the questionnaires for different modes of data collection and technologies. The plenary lectures provided an introduction to each topic, a basic theoretical background and practical advises. Copies of the power point presentations of my lectures formed the basic 'tool chest' provided to the participants. Also all participants received a copy of the excellent introductory brochure by Fritz Scheuren (2004). What is a survey.

This book goes beyond the basic course material and gives a general introduction and overview of data collection and data quality and the influence of new technology on these. For those who attended the seminar it will provide a thorough background to the practical tools that were discussed during the seminar. But this book is not restricted to attendees of the EUSTAT 2004 seminar. It can be read as a 'stand-alone' introduction into data collection methods and provides a non-technical introduction and overview to major issues.

The book starts with an overview of basic principles of quality, discussing the cornerstones of survey quality. This is followed by an overview of data collection methods, in which both practical and theoretical differences between methods are discussed. Special attention is given to advantages and disadvantages of each method, and how to take these into account and optimize the data collection procedures for the chosen purpose of a specific survey. The book ends with a systematic overview of new technologies in data collection methods and its influence on data quality. An extensive list of references is added to this book. Also each chapter ends with suggestions for more in depth reading.

As preparation to reading this book, I strongly advise to read Fritz Scheuren (2004). "What is a survey." of the American Statistical Association (freely available at www.amstat.org). For those who need practical advises on writing and testing questions, I suggest F.J Fowler (1995). Improving Survey Questions: Design and Evaluation. Thousands Oaks: Sage Applied Social Research Methods Series Vol 38 (ISBN 0-8039-4583-3).

Finally, I would like to thank EUSTAT for inviting me to teach a seminar on new technologies in data collection, questionnaire design, and quality, and for the excellent organization of the seminar. Special thanks go to the participants of the seminar for their enthusiasm and the stimulating discussions. It was a privilege and great pleasure to teach at EUSTAT .

Amsterdam, January 2005
Edith Desiree de Leeuw

2. Survey Quality

2.1. What is Quality

One of the most general definitions of quality is "fitness for use." This definition was coined by Juran and Gryna in their 1980's book on quality planning and analysis, and has been widely quoted since. How this general definition is further specified depends on the product that is being evaluated and the user. An example, quality can be focusing on construction, on making sturdy and safe furniture and on testing it. Like Ikea, the Swedish furniture chain, that advertised in its catalogues with production quality and gave examples on how a couch was tested on sturdiness.

In Survey Statistics the focus has been on 'accuracy', on reducing the mean squared error or MSE (Kruskal, 1991). The quality indicator is then the MSE: the sum of all squared variable errors and all squared systematic errors. In other words, the mean squared error is the sum of the variance plus the squared bias (Groves, 1991; 1989). However, quality of survey data has more aspects that should be taken into account. And in official statistics the view on quality has gradually changed to encompass a wider set of attributes, as the leadership group (LEG) on quality for National Statistical Institutes (Lyberg et al, 1991) emphasized in their summary report. Accuracy is no longer the sole measure of quality.

What are then important aspects of quality for statistical organizations? Biemer & Lyberg (2003) in their handbook on quality apply the concept of 'fitness for use' to the survey process. This leads to the following quality requirements for survey data: 'accuracy' as defined by the mean squared error, 'timeliness' as defined by available at the time it is needed, and 'accessibility', that is the data should be accessible to those for whom the survey was conducted. Eurostat (2000) even distinguishes seven distinct dimensions of statistical quality. These are summarized below

2.1.1. Eurostat's quality dimensions

The first component is *relevance* of the statistical concept. A statistical product is relevant if it meets user's needs. This implies that user's needs must be established at the start. The concept of relevance is closely related to validity of measurement. In other words, do we measure what we want to measure. Did we correctly translate the substantive research question into a statistical question. If not, we have made a specification error, and the statistical product does not meet the need of the user.

The second component is the well-known *accuracy* of the estimate.

The third component is *timeliness*. The results should be available on time as punctuality in dissemination is important for many users. This implies that production time and speed are very important. Also, the frequency of measurement may be important. For key indices timeliness also means recently measured, as important statistical figures should not be too old.

The fourth component is *accessibility* of information. This means more than simple accessibility. The results should not only be easily accessible, they should also be available for users in a form that is suitable to their goals!

The fifth component is *comparability*. It should be possible to make reliable comparisons across time and across space. Comparability is extremely important in cross-cultural and cross-national studies, and comparability is a prerequisite for harmonized statistics.

The sixth component is *coherence of statistics*. Like comparability coherence focuses on applied statistical concepts and definitions, and potential differences. But coherence concentrates on the *joint* use of *statistics* that are produced for different primary purposes. When statistics originate from one single source, they are coherent if they can be consistently combined in a more complex way. Statistics originating from different sources, and in particular from studies of different periodicities, are said to be coherent if they are based on common definitions, classifications and methodological standards.

The seventh and last component is *completeness*. The domains for which statistics are available should reflect the need and priorities expressed by users as a collective.

2.2. Quality in Surveys

In the above quality was discussed in the context of official statistics and national and international official statistical institutes. But, what is quality in the general context of surveys? As early as 1944, Deming already stressed that absolute quality or accuracy is a mythical concept and that it is more profitable to speak of tolerance bands or limits of likely error. He further pointed out that allowable limits must vary from case to case, depending on the resources available and the precision needed for a particular use of the data. Here we have an early reference to the needs of a survey designer to take into account both the available resources and the purpose of the survey: quality as "fitness for use".

Deming (1944) also gives an early warning of the complexity of the task facing the survey designer. He lists thirteen factors that affect the ultimate usefulness of a survey. Among those are the relatively well understood effects of sampling variability, but also the more difficult to measure effects. Deming incorporates effects of the interviewer, method of data collection, nonresponse, questionnaire imperfections, processing errors and errors of interpretation.

Hansen, Hurwitz, and Madow (1953) and Hansen, Hurwitz, and Bershad (1961), defined accuracy further in terms of survey statistics. Their model of survey errors centered on the concept of the *individual true value* of a given variable for each population element. This model differentiates between variable error and systematic bias and offers a concept of *total error*. Kish (1965) finally refers to the total error as the root mean squared error and proposes it as a replacement for the simpler but incomplete concept of the standard error. He also sets out a classification of sources of bias in survey data, differentiating between errors of nonobservation (e.g., nonresponse) and observation (e.g., in data collection and processing). This is still the basis of survey error models of today.

Recent discussion of survey errors has taken place in a variety of different contexts, such as sampling statistics, psychology, economics, often using a different terminology and framework in each context. In his book survey errors and survey costs Groves (1989) attempts

to reconcile these different perspectives and reduce the communication problems among disciplines through a generalized classification scheme developed from Kish's taxonomy. At the core is the concept of mean squared error: the sum of error variance (squared variable errors) and of squared bias. Variable (random) errors are those that take on different values over replications on the same sample, while bias (nonrandom errors) is made up of the components of error viewed to be constant over replications. Both types of error, that is random error and systematic error, can be divided into errors of nonobservation and errors of observation. Errors of nonobservation are sampling errors, noncoverage errors, and nonresponse errors. Errors of observation are also referred to as response or measurement errors. These can arise from the respondent, the interviewer, the instrument, and the mode of data collection.

Processing or coding errors are not included in Groves's system, but these are important when taking the total survey process into account (cf. Deming, 1982). Therefore, Biemer & Lyberg (2003) distinguish two broad categories of errors: (1) sampling error and (2) nonsampling error. Sampling error is due to selecting a sample instead of studying the whole population. Nonsampling errors are due to mistakes and/or system deficiencies, and include all errors that can be made during data collection and data processing, such as coverage, nonresponse, measurement, and coding error.

In a total quality approach, costs should be taken into account too (Lyberg et al, 1991). In doing this it is important not to confuse 'cost efficiency' with 'low cost'. Only by formally assessing the costs of alternative methods, while jointly measuring quality can the 'best' method be identified (Groves, 1989).

2.3. Cornerstones of Survey Data Quality

In the above the different dimensions of survey quality were discussed. How should one incorporate this when designing a quality survey? A good metaphor is constructing a house. When building a house, one starts with carefully preparing the ground and placing the corner stones. This is the foundation on which the whole structure will rest. If the foundation is not designed with care, the house will collapse or sink in the unsafe underground.

When designing a survey, one should also start by laying a sound foundation of research. One starts with preparing the underground by a detailed specification of the concepts to be measured. Then these clearly specified concepts have to be translated, or in technical terms ‘operationalized’ into measurable variables. Avoiding and reducing *specification errors* provides the safe underground needed. On this safe fundament we place the four cornerstones of survey research: coverage, sampling, response, and measurement (Salant & Dillman, 1994). Only when these cornerstones are solid, high quality data are collected, which can then be used in further processing and analysis.

2.3.1. Breaking the ground: Specification of Research Question

The first step in the survey process is to determine the research objectives. The researcher together with the client, have to agree on a well-defined set of research objectives. These are then translated into a set of key research questions. For each research question one or more survey questions are then formulated, depending on the goal of the study. For example, in a

general study of the population one or two general questions about well-being are enough to give a global indication of well-being. In a specific study of the influence of the social networks on feelings of well-being among the elderly, far more detail is needed and a series of questions has to be asked, each question measuring a specific aspect of well-being. For two examples see boxes below.

Example General Well-being Question (Hox, 1986)

Taking all things together, how satisfied or dissatisfied are you with life in general?

- ❑ VERY DISSATISFIED
- ❑ DISSATISFIED
- ❑ NEITHER DISSATISFIED, NOR SATISFIED
- ❑ SATISFIED
- ❑ VERY SATISFIED

Examples General + Specific Well-being Questions (Hox, 1986)

Taking all things together, how satisfied or dissatisfied are you with *life in general*?

- ❑ VERY DISSATISFIED
- ❑ DISSATISFIED
- ❑ NEITHER DISSATISFIED, NOR SATISFIED
- ❑ SATISFIED
- ❑ VERY SATISFIED

Taking all things together, how satisfied or dissatisfied are you with *the home in which you live*?

- ❑ VERY DISSATISFIED
- ❑ DISSATISFIED
- ❑ NEITHER DISSATISFIED, NOR SATISFIED
- ❑ SATISFIED
- ❑ VERY SATISFIED

Taking all things together, how satisfied or dissatisfied are you with *your health*?

Taking all things together, how satisfied or dissatisfied are you with *your social contacts*?

Specification errors occur when a final ***survey*** question that is posed to a respondent, fails to ask the respondent about what is essential in order to answer the ***research*** question. In other words, the concept implied in the survey question differs from the concept that should be measured; it differs from the intended concept. As a result, the wrong parameter is estimated and the research objective is not met.

A clear example of a specification error is given by Biemer & Lyberg (2003). The intended concept to be measured was " the value of a parcel of land if it were sold on a fair market today". A potential operationalization in a survey question would be 'For what price would you sell this parcel of land?' Closer inspection of this question reveals that this question asks

what the parcel of land is (subjectively) worth to the farmer. Perhaps so much that s/he would never sell it!

There are several ways in which one can investigate whether specification errors occur. First of all, the questionnaire outline and the concept questionnaire should always be thoroughly discussed with the client and explicit checks should be made if the questions in the questionnaire reflect the study objectives.

In the next step the concept questionnaire should be pre-tested with a small group of real respondent, using so called 'cognitive lab methods' (cf. Presser et al, 2004). These are qualitative techniques to investigate whether errors occur in the question-answer process. The first step in the question answer process is 'understanding the question.' Therefore, the first thing that is investigated in a pretest is if the respondents understand the question and the words used in the question as intended by the researcher. Usually questions are adapted and/or reformulated, based on the results of questionnaire pre-tests.

When a question is reformulated, there is the danger of changing the intended meaning. Again a specification error can occur. Therefore, both the results of the pretests and the final adapted questionnaire should again be thoroughly discussed with the client.

2.3.2. Placing the Corner Stones: Coverage, Sampling, Non-response, and Measurement

Sampling and Sampling Error

In surveys, usually a sample is investigated instead of the whole population. *Sampling error* occurs because only a sample of the population is investigated instead of the whole population. Based on the values for the variables in the sample, the value for the population is estimated using statistical theory. When simple random sampling is used, standard statistical techniques can be used. However, when more complicated sampling schemes are used, such as cluster sampling, the standard statistical techniques do not provide accurate p-values and confidence intervals and more complicated statistical techniques should be used. Multilevel analysis (Hox, 2002) is a very flexible technique to accommodate these problems.

Coverage and Coverage Error

When doing a survey one has an intended population in mind: the target population. To draw a sample from the target population, a 'sample frame' is needed. This can be a list of target population members, for instance, a list of all members of a certain organization, or the register of all inhabitants of a certain city. But it may also be a virtual list, or an algorithm, such as in area probability sampling or in Random Digit Dialing (RDD) sampling. In the later, random telephone numbers are generated using an algorithm that conforms to properties of valid telephone numbers in the country that is being investigated. Frame *coverage errors* occur when there is a mismatch between the sampling frame and the target population. In other words when there is no one-to-one correspondence between the units in the frame and the units in target population.

The most common form of coverage error is undercoverage, that is, not all units of the target population are included in the sampling frame. A clear example of undercoverage is persons with an unlisted phone number when the sampling frame is the telephone book. Another form

of coverage error is overcoverage; here a unit from the target population appears more than one time in the sampling frame. Duplications like this can occur when a sampling frame results from the combination of several lists. For example, on one list a woman is listed under her maiden name, and on a second list under her married name. If these lists are combined, the same person is listed under two different entries. A third type of coverage error is caused by erroneous inclusions in the frame. For example, a business number is included on a list with household phone numbers.

Response and Nonresponse Error

Nonresponse is the inability to obtain data for all sampled units on all questions. There are two types of nonresponse in surveys: unit nonresponse and item nonresponse. Unit nonresponse is the failure to obtain any information from an eligible sample unit. For instance, through noncontact or refusal. (For a detailed discussion of unit nonresponse and error, see Couper & De Leeuw, 2003). Item nonresponse or item missing data refers to the failure to obtain information for one or more questions in a survey, given that the other questions are completed (see also De Leeuw, Hox, Huisman, 2003).

In quality surveys, nonresponse figures should be clearly reported. This often takes the form of a response rate figure. When reporting response rates it is extremely important to clearly state the way the response rate was calculated. For details of response rate calculation and a description of sources of nonresponse, see the brochure Standard definitions of the American Association for Public Opinion Research (AAPOR). A regularly updated version can be found on the website of AAPOR (www.aapor.org).

Nonresponse error is a function of the nonresponse rate and the differences between respondents and nonrespondents. If nonresponse is the result of a pure chance process, in other words if nonresponse is completely at random, there is no real problem. Of course, the realized sample is somewhat smaller, resulting in slightly larger confidence intervals around estimators. But the conclusions will not be biased due to the nonresponse. Only when respondents and nonrespondents differ from each other on the variables of interest in the study, there will be a serious nonresponse problem. The nonresponse is then *selective* nonresponse and certain groups may be underrepresented. In the worst-case scenario, there is a substantial association between the nonresponse and an important variable of the study causing biased results. A classic example comes from mobility studies: people who travel a lot are more difficult to contact for an interview on mobility than people who travel rarely. This selective nonresponse caused by specific noncontacts, may then lead to an under-estimate of mobility.

Measurement and Measurement Error

Measurement error is often also called error of observation. Measurement errors are associated with the data collection process itself. They occur when we ‘de facto’ collect the data for our survey. There are three main sources of measurement error: the questionnaire, the respondent, and the method of data collection. In those cases when an interviewer is used for data collection, the interviewer is the fourth source of error.

A good designed and well-tested questionnaire is the basis for reducing measurement error. The question should be clear, and all respondents should understand the terms used in the same way. With a closed-ended question, the response categories should be well defined, and

exhaustive. When a question is not clear, or when the answer-categories are not clearly defined, respondents will make errors while answering the question, or they do not know what to answer. Interviewers will try to help out, but in doing this they can make errors too, and introduce additional interviewer error (Fowler, 1991). Therefore, improving the questionnaire is a good start to improve the total survey quality. It should be emphasized that even carefully designed questionnaires still may contain errors, and that a questionnaire should always be evaluated and pre-tested before it may be used in a survey. I will not go into details on question writing and questionnaire testing. A good handbook on this topic is the one by Fowler (1995).

Respondents can be a source of error in their own right, when they provide the researcher with incorrect information. This may be unintentionally, for instance when a respondent does not understand the question, or a respondent has difficulty to remember an event. But a respondent can also give incorrect information on purpose; for instance when sensitive questions are asked. Measurement errors that originate from the respondent are beyond the control of the researcher. A researcher can only try to minimize respondent errors by making the respondent's task as easy and as pleasant as possible. Or in other words, by writing clear questions that respondents want to answer. Besides a good questionnaire, a careful choice of mode of data collection helps to reduce respondent error.

The method of data collection can be a third source of measurement error. For instance, in a telephone interview respondents have to rely on auditive cues only: they only hear the question and the response categories. This can cause problems when a long list of potential answers has to be presented. For a more detailed discussion, see chapter 3.

Finally, when an interview is conducted, the interviewer can be a source of error too. Interviewers may misinterpret a question, may make errors in administering a questionnaire, or in registering the answers. When posing the question, interviewers unintentionally may change the meaning of the questionnaire. By giving additional information or explaining a misunderstood word, they may inappropriately influence a respondent. Even the way they look and dress may influence a respondent in a face-to-face interview. Selecting and training interviewers carefully will help in reducing interviewer related errors. For a short introduction into training issues, see Czaja & Blair (1996). Interviewers can make genuine mistakes, but they also may intentionally cheat. Interviewers have been known to falsify data, or skip questions to shorten tedious interviews. Monitoring interviewer helps to reduce this. Having a quality controller listening in on telephone interviewers is a widely used method. In face-to-face interviews, recordings can be made and selected tapes can be checked afterwards. Special verification contacts or even re-interviews are also used to evaluate interviewer performance in large-scale face-to-face surveys. For an overview of interviewer evaluation methods, see Biemer & Lyberg (2003, chapter 5).

2.4. Quality Control

Central in modern quality control is the principle of *Total Quality Management* or TQM. The origin of TQM is statistics, and the first principles were formulated by Deming (1944, 1982). However the first major applications were in industry; for instance, Japanese electronic industries were early adaptors of TQM-principles. Statistical quality has been always important in the statistical and survey world, but it took until the end of the twentieth century for TQM to be completely established in survey methodology (see for instance Lyberg et al,

1997, section D). Total Quality Management implies that one aims at continuous quality improvement for all stages in the survey process. The stages of the survey process are summarized in Table 2.1 below

Table 2.1 Total Survey Process

The main stages in surveys for official statistics are listed

1. Determining the Research Objective
2. Defining target population and Constructing sampling frame
3. Selection of sample
4. Compilation of questionnaire
    a. Construction
    b. Pretesting
    c. Programming (when computer assisted methods are used)
5. Data Collection
6. Processing
    a. Coding
    b. Checking
    c. Editing
7. Production final data set
    a. Imputation
    b. Nonresponse Adjustment
8. Estimation and Analysis
9. Publication of results
10. Dissemination of statistical information
11. Archiving

In each stage of the survey process known *Current Best Methods* (CBM) should be applied and in each stage quality evaluation should take place. If necessary, the CBM's in use should be updated or adapted to cope with weaknesses and to improve quality. This is a continuing process, as the current best methods of today, may be obsolete tomorrow. Therefore, research agencies have to invest in CBM and in quality improvement.

A good tool in quality evaluation is the survey *quality profile*. This is a report that provides a comprehensive picture of the whole survey process for a specific survey and specifies every potential error source (e.g., sampling, coverage, nonresponse, specification error, measurement error, data processing error). This means that for every survey a summary is provided of what is known about that survey for all error sources. This leads to recommendations for improvement. These recommendations for reducing errors are clearly specified for specific areas. In the next step, areas in the survey process are identified where knowledge about errors is deficient, and procedures are outlined for further methodological research into these areas.

After identifying errors follows coping with errors. Again, a total survey error approach should be used. Each error source of importance has to be identified and the relative

importance of each source has to be assessed. The resulting weights can be different for different surveys, as the specific survey objectives determine the relative weights for each error source. Finally the costs should be taken into account, both the potential changes in survey costs and control costs. The final goal is to find that combination of important design parameters that gives the highest quality for the available budget. Thereby making it possible to allocate available resources according to a total survey error minimization strategy.

Finally, essential for a successful implementation of Total Quality Management is the documentation and dissemination of knowledge. Research agencies should carefully document their CBM-procedures and make sure that the CBM-procedures are well-known in the organization. Potential tools are easily accessible publications of quality guidelines and in-house seminars, workshops and courses. Good examples of quality guidelines for official statistics available in English are, among others, the quality guidelines of Statistics Finland (Handbook 43b), and the AAPOR guidelines for designing surveys in public opinion research (available at www.AAPOR.org). A good source is the web page of the Australian Statistical Office, which has a section on international practice with helpful links (see also section 2.5).

In Sum: The goal of quality control is to assess, control, and/or compensate for survey errors. Necessary is to document the procedures used for pre-survey evaluation, for quality assurance during the survey and for post survey evaluation. One should always keep in mind that there are more than one sources of error, and that one has to compromise and choose when attempting to reduce total survey error. Listing the stages of the survey and describing potential sources of error in a survey plan and identifying the combination of most important design parameters for the survey goal makes it easier to allocate the research funds wisely. For more details see Biemer & Lyberg, 2003, especially chapter 8 and 10.

2.5. Suggested Readings and Websites

Handbooks:

On survey quality

P. P. Biemer & L. E. Lyberg (2003) Introduction to Survey Quality, New York: Wiley

R.M. Groves (1989). Survey errors and survey costs, New York, Wiley.

Statistics Finland (2002). Quality Guidelines for Official Statistics, Handbooks 43b: Helsinki: Statistics Finland also available at http://www.stat.fi/tk/tt/laatuatilastoissa/guidelines.pdf; (www.stat.fi/qualityguidelines)

On question writing and testing

F.J. Fowler (1995). Improving Survey Questions: Design and Evaluation. Thousands Oaks: Sage Applied Social Research Methods Series Vol 38

Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E, Martin, J., Singer, E. (2004). Methods for testing and evaluating survey questionnaires. New York: Wiley.

On practical aspects of surveys

R. Czaja & J. Blair (1996). Designing surveys: A guide to decisions and procedures. Thousand Oaks: Sage (Pine Forge Press series in research methods and statistics). An updated and revised version will be published in 2005.

Web sites:

On general survey methods:

Webpage Australian Statistical Office at http://www.sch.abs.gov.au/

This website has an excellent section on international practices with links to best practice quality standards.

European quality criteria (Eurostat):

http://www.unece.org/stats/documents/2000/11/metis/crp.3.e.pdf

Response rate standard definitions:

http://www.aapor.org/pdfs/standarddefs2004.pdf

3. Modes of Data Collection

3.1. Main Modes of Data Collection

In survey research traditionally three major methods of data collection can be distinguished: the face-to-face interviews, the telephone interview, and the postal (mail) questionnaire (De Leeuw, 1992). The *face-to-face interview* is the mode in which an interviewer administers a structured or partly structured questionnaire to a respondent within a limited period of time and in the presence (usually at the home) of the respondent. The face-to-face interview is the oldest mode of data collection and was the sole serious method of data collection until the late seventies of the last century. The telephone interview is a younger method of data collection. Telephone surveys became the predominant mode in the USA around 1980. In a *telephone interview* the interviewer administers the questions (from a structured questionnaire and within a limited period of time) via a telephone. Telephone interviewing is often centralized; i.e., all interviewers work from a central location under direct supervision of a field manager or a quality controller. Mail surveys are almost as old as face-to-face surveys, and have been through a revival in last thirty years. When a *mail questionnaire* is used, a respondent receives a structured questionnaire and an introductory letter by mail, answers the questions in her/his own time without any assistance (from the researcher or her/his representative) except for any written instructions in the questionnaire or in the accompanying letter, and finally sends the questionnaire back.

Other modes of data collection are diary surveys, direct observation, and the use of administrative records. Diary surveys, such as time budget studies, are a self-administered form of data collection that asks a high involment of the respondent. Direct observation, needs trained observers but does not place any burden on 'respondents'. Administrative data records can in special cases be used to collect data. Usually, only government organizations with a special permit are allowed to combine and link administrative records.

It should be noted that computer assisted procedures for all data collection techniques have been developed in the last thirty years, of which CATI (computer assisted telephone interviewing) is the oldest form. Besides CATI, these procedures include CAPI (computer assisted personal interviewing), and various forms of computer assisted self-administered questionnaires. Recently much attention has been paid to a special form of the computer assisted self-administered questionnaire: the Internet or web survey (see for instance, Dillman, 2001, Couper, 2000; Czaja & Blair, 2005). For a further discussion of computer assisted data collection see chapter 4 of this book.

In the remainder of this chapter I will concentrate on the three main modes: face-to-face, telephone and mail survey. For a concise overview of other methods, see Biemer & Lyberg, 2003.

3.2. Practical Advantages and Disadvantages

3.2.1. Coverage: Population of interest and possibility of sample control

When one is interested in studying the general population, the face-to-face survey has the greatest potential. Sophisticated sampling designs for face to face surveys have been developed,

which do *not* require a detailed sampling frame or a list of persons or households (cf. Kish, 1965). For instance, area probability sampling can be used to select geographically defined units (e.g., streets or blocks of houses) as primary units and households within these areas. Therefore, a main advantage of face-to-face interviews is its potential for a high coverage of the intended population. Elaborate techniques based on household listings (e.g., inventories of all household members derived by an interviewer) can then be used to randomly select one respondent from those eligible in a household; for an overview, see Gaziano, 2005). Face-to-face interviewing has the highest potential regarding coverage and sampling, but it can be very costly, especially if the country is large and sparsely populated. Cluster sampling may be needed, and if the sample dispersion is very high telephone surveys are often employed.

Telephone interviews are feasible if telephone coverage is high, in other words if the non-telephone part of the population can be ignored (cf. Biemer & Lyberg, 2003). To be sure that persons with unlisted telephones are also included, one can employ random digit dialing. Random digit dialing techniques, which are based on the sampling frame of all possible telephone numbers, make it possible to use telephone interviews in investigations of the general population. A new challenge to telephone survey coverage is the increasing popularity of mobile (cell) phones. If mobile phones are additional to fixed landline phones this will not pose a major problem for coverage. But, there is evidence that certain groups (e.g. the young, lower income, urban, more mobile) are over represented in the mobile phone only proportion of the population. For an overview see Nathan (2001). In telephone interviews, as in face-to-face interviews, the Kish procedure based on a complete household listing can in theory be used to select respondents within a household. However asking for a complete household listing over the phone, is a rather complex and time consuming procedure and increases the risk of break-offs. A good alternative for the Kish procedure is the last birthday method. In the last birthday method, the interviewer asks to speak with that household member who most recently had a birthday.

Mail surveys require an explicit sampling frame of names and addresses. Often, telephone directories are used for mail surveys of the general population. Using the telephone directory as a sampling frame has the drawback that people without a telephone and people with an unlisted telephone cannot be reached. The reason for the frequent use of the telephone directory as sampling frame is the relative ease and the low costs associated with this method. A drawback of mail surveys is the limited control the researcher has over the choice of the specific individual within a household who in fact completes the survey. There is no interviewer available to apply respondent selection techniques within in a household and all instructions for respondent selection have to be included in the accompanying letter. As a consequence only simple procedures as the male/female/youngest/oldest alternation or the last birthday method can be successfully used. The male/female/youngest/oldest alteration asks in a random 25% of the accompanying letters for the youngest female in the household to fill in the questionnaire, in 25% of the letters the youngest male is requested to fill in the questionnaire, et cetera.

In Internet or web surveys, coverage is still a major problem when surveying the general population (Couper, 2000). Like in mail surveys the control of the interview situation is low.

When a complete list of the individual members of the target population is available, which can be the case in surveys of special groups or in countries with good administrative records, a random sample of the target population can be drawn regardless of the data collection method used. In that case, coverage and sampling will not be a decisive issue in the choice of data collection.

3.2.2. Nonresponse

Survey nonresponse is the failure to obtain measurements on sampled units. Nonresponse can be distinguished from another error of nonobservation, coverage error (discussed above), by the fact that nonrespondent units are selected into the sample, but not measured, whereas noncovered units have no chance of being selected in any sample (e.g., no known address, no telephone number) and thus cannot be measured. There are two major sources of nonresponse: noncontact in which no request for cooperation can be made, and explicit refusal. A third source is 'incapacity' to cooperate. Examples of method-specific incapacities to answer are illiteracy in mail and web surveys, and deafness and language problems in telephone and face-to-face surveys (see also Dillman, Eltinghe, Groves, & Little, 2002). For a detailed discussion of sources of nonresponse and response rate calculation see the standard definitions of the American Association for Public Opinion Research (AAPOR, 2004). A regularly updated version can be found on the AAPOR website (www.aapor.org).

Response rates can be influenced by many factors: the topic of the questionnaire, the length of the questionnaire, the survey organization, the number of callbacks or the number of reminders, and other design features. In this section I will only discuss so called "cold" surveys (i.e., surveys for which a fresh sample is drawn). Surveys that use a panel design or a 'respondent pool' or 'access panel' of respondents who are willing to participate in on-going research, will in general have a much higher response rate than cold surveys. The reason for this higher response is that the hard-core nonrespondents have already been filtered out in panels in the acquisition stage.

In general, nonresponse has been increasing over time. For instance, De Leeuw and De Heer (2001) showed that for a large variety of surveys in official statistics, response rates have been declining internationally. They analysed data from 16 different countries over the period 1980-1998, and found an increase in both noncontacts and refusals over the years.

Face-to-face surveys tend to obtain higher response rates than *comparable* telephone surveys. Mail surveys tend to have a lower response rate than comparable face-to-face and telephone surveys. Goyder (1987) published one of the first systematic overviews on differences in nonresponse among modes. He collected data on 385 mail surveys, 112 face-to-face surveys and 53 telephone surveys in the U.S.A. and Canada between 1930 and 1980. On *average* the response rate for the face-to-face interview was 67.3%, for the telephone interview 60.2%, and for the mailed questionnaire 58.4%. Goyder (1987) also notes a pronounced increase in nonresponse for the face-to-face interview over the years, while the nonresponse for mail surveys remains stable. Hox and De Leeuw (1994) came to similar conclusions. Their meta-analysis summarized the results of 45 studies that explicitly compared the response obtained in mail, telephone, and face-to-face surveys. The data for these 45 mode comparisons were collected in several countries in Europe, in the USA, and in Canada. Again, on *average* face-to-face interviews produced the highest response (70.3%), telephone interviews the next highest (67.2%), and mail surveys the lowest (61.3%). The trend remarked upon by Goyder (1987), is clearly visible in the data of Hox & De Leeuw (1994). Both the face-to-face and telephone surveys show a decrease in response over time, while the response of mail surveys remains stable over time. Similar results were found in Germany for the time period 1960-1995 (Bretschneider & Schumacher, 1996). It should be noted that all figures cited were based on official (government) surveys and on semi-official and academic surveys at the end of the twentieth century. Response figures for commercial and market research surveys are in general much lower.

Systematic overviews of response rates in Internet surveys are not yet available. For non-scientific 'pop-up' web surveys, where an invitation to complete a survey pops-up on a web portal, the response cannot be determined. The reason why the response rate cannot be computed for pop-up web surveys is that the total number of eligible respondents is not known and the population not well-defined. When a good sampling frame is available and a sample is drawn, response rates for web surveys can be computed. The first results for these scientific probability based web surveys are promising (Vehovar, Batagelj, Lozar Manfreda, & Zaletel, 2002). Empirical comparisons between e-mail and paper mail surveys of the same population indicate that response rates on e-mail surveys are lower than for comparable paper mail surveys (Couper, 2000)

3.2.3. Questionnaire: Question type and complexity of instrument

Face-to-face interviews are the most flexible form of data collection method. The presence of a well-trained interviewer enables the researcher to use a large variety of measurement instruments. Structured or partly structured questionnaires can be used, and respondents can be asked to sort objects or pictures. Highly complex questionnaires can be successfully implemented as a trained interviewer takes care of the navigation through the questionnaire. Also, respondents can be presented with all kinds of visual stimuli, ranging from simple response cards listing the response categories of a question, to pictures, advertisement copy or video clips. In computer-assisted face-to-face interviews (CAPI), the interviewer is guided through the (complex) questionnaire by a computer program. This lowers error rates even more and gives the interviewer more opportunities to concentrate on the interviewer-respondent interaction and the respondent tasks (see also chapter 4).

Telephone interviews are less flexible. Their major drawback is the absence of visual cues during the interview; telephone is auditive only. No response cards with lists of response categories are available; the interviewer reads aloud the question and the available response categories and the respondent has to rely solely on memory. Therefore, only questions with a limited number of response categories can be used. This has led to the development of special question formats for questions with seven or more response categories (e.g., the two step or unfolding procedure), and verbal alternatives for graphically presented questions like the political "thermometer" (cf. Dillman, 1978, chapter 6). In general, question must be short and easily understandable over the phone. However, just as in face-to-face interviews, well-trained interviewers are an advantage. In telephone surveys the interviewer can assist respondents in understanding questions, can administer questionnaires with a large number of screening questions, control the question sequence, and probe for answers on open questions. Again like in CAPI, the use of computer-assisted telephone interviewing (CATI) makes these tasks easier for the interviewer.

The absence of an interviewer makes mail surveys the least flexible data collection technique when complexity of questionnaire is considered. All questions must be presented in a fixed order, and only a limited number of simple skips and branches can be used. For routings like skips and branches special written instruction and graphical aids, such as arrows and colours, have to be provided. In a mail survey, all respondents receive the same instruction and are presented with the questions without added interviewer probing or help in individual cases. In short, a mail questionnaire must be totally self-explanatory. But, a big advantage is that visual cues can be used, and with well-developed instructions fairly complex questions and attitude scales can be used. The visual presentation of the questions makes it possible to use all types of

graphical questions (e.g., ladder, thermometer), and to use questions with seven or more response categories. Also, information booklets or product samples can be sent by mail with an accompanying questionnaire for their evaluation. Another advantage is that mail surveys can be completed when and where the respondent wants. A respondent may consult records if needed, which may improve accuracy, and the greater privacy is an advantage with sensitive topics.

Internet surveys share the advantages of mail surveys regarding visual aids. Also, as in mail surveys, the respondent is in charge and the situation may offer more privacy. Because an interview program determines the order of the questions, more complex questionnaires can be used than in a paper mail survey. In this sense (complexity of questionnaire structure) an Internet or web survey is equivalent to an interview survey. But, Internet also has a drawback, it is a more perfunctory medium and people often just pay a flying visit. Respondents may have a stronger tendency to satisfy and give top-of-the head answers (cf. Schwarz et al, 1991).

3.2.4. Questionnaire: Questionnaire length and duration of interview

Regarding the duration of the interview and the amount of questions asked the face-to-face interview again has the most potential. Face-to-face interviews can last longer than either telephone or mail surveys. It takes a highly assertive respondent to end an overly long face to face interview, while this is much easier in a telephone and especially in a mail survey. Terminating a web surveys is easiest of all, a break-off is just one mouse-click away! As a rule, successful telephone surveys can be conducted with an average length of twenty to thirty minutes. Longer telephone interviews will lead to either a somewhat higher nonresponse rate or a higher probability of premature termination of the interview. Still, successful telephone interviews have been reported which took over 50 minutes. Both Heberlein & Baumgartner (1978) and Goyder (1982) found a small negative effect of length of questionnaire on the response rates of mail surveys. According to Dillman (1978, p. 55) mail questionnaires up to 12 pages, which contain less than 125 items, can be used without adverse effects on the response. Internet surveys must be relatively short; 10-15 minutes is already a long time for an Internet survey (cf. Czaja & Blair, 2005).

3.2.5. Resources available: Completion time, organization, personnel and costs

Each data collection technique requires that certain organizational requirements are met. In general, Internet and telephone surveys are the fastest to complete for a survey organization. Mail surveys are usually locked into a definite time interval of mailing dates with rigidly scheduled follow ups, and geographically dispersed face-to-face interviews take the longest time. When speed of completion is really important and data are needed very quickly, telephone and Internet surveys are best. If the data are needed in a couple of weeks, mail surveys are also feasible (Biemer & Lyberg, 2003; Czaja & Blair, 2005).

Dillman (1978, p. 68) gives an example in which a survey unit of 15 telephones can complete roughly 3000 interviews during the 8 weeks it takes to do a complete TDM mail survey. Only if the telephone unit is smaller than 15 interviewers, or the number of needed completed interviews larger than 3000, a mail survey will be faster.

The implementation of a successful, large scale, face-to-face survey demands most from an organization and its personnel. Interviewers have to be trained, not only in standard interview techniques, but also in how to implement sampling and respondent selection rules, and in how to solve various problems that can arise when they are alone in the field. In addition, an extensive supervisory network is needed to maintain quality control. Finally, an administrative manager is needed to make sure that new addresses and interview material are mailed to the interviewers on a regular base.

The personnel requirements for a telephone survey are less demanding. Because of the centralized setting, fewer highly trained supervisors are needed. Interviewers should, of course, be well trained in standard interview techniques. But, because of the close supervision the variety of skills needed is less. The majority of the interviewers no longer have to be prepared for every possible emergency and can concentrate on standard, but good quality interviewing. Difficult respondents or problem cases can be dealt with by the available supervisor or can be allocated to a specially trained interviewer.

Organizational and personnel requirements for a mail survey are even less demanding. Most of the workers are not required to deal directly with respondents, and the necessary skills are mainly generalized clerical skills (e.g., typing, sorting, response administration, and correspondence processing). Of course, a trained person must be available to deal with requests for information, questions, and refusals of respondents. Finally, the number of different persons needed to conduct a mail survey is far less than that required for face to face or telephone surveys with equal sample sizes. For instance, one person can single-handedly successfully complete a TDM mail survey of a sample of 1000 persons in the prescribed 8 week TDM schedule (De Leeuw, 1992). However, to design and implement an Internet survey highly skilled and specialized personnel is needed. To design a successful Internet survey both technical knowledge is needed (e.g., operating systems, browsers, etc) and knowledge on usability and visual design (cf. Czaja & Blair, 2005; Couper, 2000; Dillman, 2000).

Requirements for the organization and personnel do influence the cost of data collection. Mail and Internet surveys have relatively low costs, and may be the only modes affordable in certain situations. Telephone surveys are less expensive than face-to-face modes, especially in widely geographically dispersed surveys. When interviewer-assistance is essential, but the survey is a large national or international study, telephone surveys are the only option.

3.3. Data Quality

3.3.1. Mail, Telephone, and Face-to-face Surveys

The influence of data collection method on data quality has been extensively studied for face-to-face interviews, telephone surveys, and self-administered mail questionnaires. De Leeuw (1992) performed a meta-analysis of known empirical comparisons. The resulting overview showed clear differences between methods, suggesting a dichotomy of survey modes in modes with and modes without an interviewer. Comparing mail surveys with both telephone and face-to-face interviews, De Leeuw found that it is somewhat harder to have people answer questions in mail surveys. Both the overall non-response and the item nonresponse are higher in self-administered questionnaires when compared with interviews. However, when questions are answered, the

resulting data tend to be of better quality. Especially with more sensitive questions, mail surveys performed better (less social desirability, more reporting of sensitive behaviour like drinking, less item nonresponse on income questions).

When face-to-face and telephone surveys were compared small difference in quality were discovered. Face-to-face interviews resulted in data with slightly less item nonresponse. No differences were found concerning response validity (record checks) and social desirability. In general, similar conclusions can be drawn from *well-conducted* face to face and telephone interview surveys (De Leeuw, 1992; De Leeuw & Van der Zouwen, 1988).

In a follow-up study De Leeuw (1992) investigated additional aspects of data quality, such as consistency and reliability of answers, response tendencies, and responses to open questions. Again, the main differences were between the mail survey on the one hand and the two interview surveys on the other hand. The mail survey resulted in more reliable and consistent responses, and less acquiescence. However, the differences are relatively minor. Regarding responses to open questions, the results are mixed. When short open questions are asked on well-defined topics, the differences are small. With more complex questions, the assistance and probing of an interviewer is necessary to get detailed answers.

In sum: When comparable surveys with equivalent questionnaires are investigated none of the data collection modes was superior on all criteria. The most pronounced differences were found with more sensitive topics. The modes with an interviewer produced more socially desirable answers and less consistent answers, but also more detailed responses to open questions. Differences between face-to-face and telephone interviews were small, with the face-to-face interview doing slightly better than the telephone.

3.3.2. Computer-assisted modes

Direct comparisons of different forms of computer-assisted methods are rare. Most of the literature is on comparisons between paper-and-pencil methods with their computer-assisted equivalent (see chapter 4). I can extrapolate the main findings on comparisons of paper-and-pencil survey methods to the new computer-aided forms of data collection methods (see also De Leeuw & Collins, 1997). For respondents in a telephone interview nothing changes when a research institute switches from paper-and-pencil telephone surveys to CATI. For the interviewers the task becomes less complex, because administrative duties have been taken over by the computer. As a result, the differences, if any, point toward a slight advantage for CATI, for instance fewer routing errors. In CAPI the computer is visible to the respondent, who might react to its presence. However, very few adverse reactions and no reduction in response rates have been reported. It seems safe to assume that the main findings concerning mode differences between telephone and face-to-face surveys are also valid for the computer-aided versions of these survey techniques. This means that with well-trained interviewers and the *same* well-constructed structured questionnaire, both CAPI and CATI will perform well and differences in data quality will be extremely small. Of course, it should be noted that CAPI has a greater potential than CATI, just as paper-and-pencil face-to-face interviews have a greater potential than paper and pencil telephone interviews. Visual stimuli may be used and more complex questions asked.

There are several forms of computer aided self-administered questionnaires or CASI (see also chapter 4). Just as in paper-and-pencil self-administered questionnaires the respondents in CASI

answer the questions in a private setting, which reduces a tendency to present themselves in a favorable light. In line with this, there is some evidence that CASI produces fewer socially desirable answers than CAPI when sensitive questions are asked. For a more detailed discussion, including context effect, see De Leeuw & Collins (1997).

3.4. Mixed-Modes Data Collection

Instead of collecting survey data with one single mode, it is also possible to use a mix of data collection methods and implement a mixed-mode design. Mixed-mode designs have been customary within longitudinal surveys for a long time. One of the main reasons for a mixed-mode approach in longitudinal studies is costs. Usually the first measurements in the base-line survey is done face-to-face, wile in the next wave less expensive methods are used, such as mail surveys. In cross-sectional studies a mixed-mode approach is often used to raise the overall response of the survey. The data collection starts with the least expensive method, for instance, a mail survey. Later on in the data collection phase the nonrespondents are approached with a more costly interview method.

Time and money restraint are the main reason for employing a mixed-mode approach in household surveys when all household members have to be surveyed. After a face-to-face interview with one household member, self-administered questionnaires are left for the other members to be mailed back. In this case, the face-to-face interview consists of both general questions about the household and specific questions for the individual household member. The self-administered questionnaire is shorter and consists only of the specific questions; the general information on the household is already known from the interview.

Another practical reason for employing a mixed-mode survey is geographical dispersion of certain groups. In densely populated areas a face-to-face interview may be feasible, but in sparsely populated this will be too costly. For instance, in the European Social Survey most countries are surveyed face-to-face, but several sparsely populated countries use telephone surveys.

Reducing errors and improving quality is another ground for mixing modes. To reduce coverage error mixed-mode designs have been used for a long time in telephone surveys and dual frame sampling schemes have been developed, combining traditional area probability sampling and face-to-face interviewing with telephone sampling and telephone interviewing (cf. Groves & Lepkowsky, 1985). To reduce social desirability bias, often a mixed-mode approach is used of a general interview plus a self-administered questionnaire for the sensitive questions. During or after the interview the interviewer hands over a questionnaire, asks the respondent to complete this questionnaire in privacy, and hand the completed form back in a sealed envelope. This mixed-mode approach results in better data quality than a single mode where all questions (sensitive and non-sensitive) are posed by an interviewer (Makkai & McAllister, 1992). The same procedure can be uses during a CAPI-interview, resulting in a CAPI-CASI mix. In this variation, the interviewer gives a short introduction and hands over the computer to the respondent.

A very recent mix is the telephone-internet combination. Several forms are feasible. For instance after a short telephone interview respondents may be asked if they want to become a member of an Internet panel, or the telephone survey is used as a screening tool for specific groups, who are asked to go to a web address and fill-out a questionnaire. Other mixes with Internet are leaving

respondents free to choose their favorite method (e.g., mail survey, telephone, internet) or follow-up the nonrespondents to an Internet survey by phone.

A prerequisite for a successful mixed-mode approach is a careful consideration of potential mode effects. In the case of sensitive questions and the use of a self-administered questionnaire during an interview, one is exploiting the mode effects to get better answers. But in other situations, the researcher needs to minimize mode effects (see also Dillman & Christian, 2003). A prime example is a longitudinal study. The goal of a longitudinal study is comparison over time: one wants to measure real change, not a change that is caused by switching modes. In this case it is important that the questions are worded in the same way in both modes, and the same question format is used. In other words, the questionnaires should be equivalent. This often means that the full potential of a mode cannot be used. For instance, when changing from a face-to-face interview to a telephone interview, one cannot use visual aids and long lists of response categories. When switching from face-to-face to mail or Internet one can use visual aids, but must realize that interviewer assistance with complex questions is not available. Still there is a common middle ground, where equivalent questionnaires of good quality can be constructed for different modes. In those cases mixed-mode approach may be the best choice!

3.5. Summary

An optimal data collection method is the best method given the research question and given certain restrictions. The basic research question defines the population under study and the types of questions that should be asked. Questions of survey ethics and privacy regulations may restrict the design. Important practical restrictions are time and funds available and other resources (skilled labour, administrative resources, experience, computer hardware and software). Within the limits of these restrictions difficult decisions have to be made concerning, for example, the acceptable level of nonresponse or the acceptable level of measurement error and the desired data quality.

It is not necessary that a researcher should restrict the survey design to one method only. Mixing methods may have advantages, and after careful consideration a combination of methods may be the best choice for a specific research question.

3.6. Suggested Readings and Websites

Handbooks:

On survey quality and data collection

P. P. Biemer & L. E. Lyberg (2003) Introduction to Survey Quality, New York: Wiley (especially chapter 5 & 6)

On practical aspects of surveys

R. Czaja & J. Blair (1996). Designing surveys: A guide to decisions and procedures. Thousand Oaks: Sage (Pine Forge Press series in research methods and statistics). An updated and revised version has been published in 2005. This second edition includes many practical aspects of Internet surveys too.

For the USA-situation

R.M. Groves, F.J. Fowler, Jr., M.P. Couper, J.M. Lepkowski, E. Singer, & R. Tourangeau (2004) Survey Methodology. New York: Wiley (especially chapter 5).

Web sites:

On general survey methods:

Webpage Australian Statistical Office at http://www.sch.abs.gov.au/

Response rate standard definitions:

http://www.aapor.org/pdfs/standarddefs2004.pdf

On Internet surveys:

http://www.websm.org/

4. New Technologies and Data Quality: A Review of the Evidence

4.1. New Technologies and Data Collection

New technologies have had a strong impact on the daily practice of data collection, and whether computer assisted data collection should be used in survey research seems now no longer an issue of debate. Computer assisted methods are replacing paper-and-pen methods at an increasing pace, and in Europe and North America many government survey organizations now increasingly employ these new methods for their surveys. Large market research organizations and academic research organizations have followed (Collins, Sykes & O'Muircheartaigh, 1998). For a historical overview of the development of computer assisted data collection see Couper and Nicholls (1998) and Nicholls and De Leeuw (1996).

Characteristic of all forms of computer assisted data collection is that questions are read from the computer screen, and that responses are entered directly in the computer, either by an interviewer or by the respondent self. An interactive program presents the questions in the proper order; in more advanced forms this order may be different for different (groups of) respondents. In chapter 3 various forms of paper-and-pen data collection methods have been discussed. For each paper-and-pen form there is now a computer-assisted equivalent available. Table 4.1 on the next page gives an overview of data collection methods and their computer-assisted equivalents. The different forms of computer assisted data collection are described below. A more detailed overview can be found in De Leeuw, Hox, and Snijkers (1998).

Forms of Computer Assisted Data Collection

Computer Assisted Telephone Interviewing (CATI) is the oldest form, and is also the most prevalent. CATI is now the dominant method for telephone surveys in market research, government organizations and universities, although paper-and-pencil methods are still being used with good results in smaller survey organizations and for short surveys. For face-to-face interviews, Computer Assisted Personal Interviewing (CAPI) is rapidly gaining in popularity and is already widely used in government statistical agencies. Large market research firms and research departments at universities are following. The later were very quick to see the potential of CAPI when surveying special populations, especially in combination with Computer Assisted Self Interviewing (CASI): a computer assisted form of the self-administered questionnaire.

Computerized self-administered data collection takes many forms. The oldest is the electronic questionnaire or electronic test, which is used in the medical and psychological sciences. In survey research, CASI is frequently used during CAPI-sessions on sensitive topics, when the interviewer hands over the computer to the respondent for a short period, but remains available for instructions and assistance. This is equivalent to the traditional procedure where an interviewer might give a paper questionnaire to a respondent to fill in privately. A very promising variant is Audio-CASI, where the respondent listens to the questions read by a computer-controlled digitized voice over a headset, and at the same time sees the question on the computer screen. This helps overcome literacy problems with special populations and

guarantees the privacy of the respondent.

Table 4.1. Taxonomy of Computer Assisted Data Collection Methods

Presented is a systematic overview of survey methods and their computer assisted equivalents. *General names*: CADAC (Computer Assisted Data Collection), CASIC (Computer Assisted Survey Information Collection, CAI (Computer Assisted Interviewing).

Data Collection Method	**Computer Assisted Form**
Interview:	
Face-to-face interview	CAPI (Computer Assisted Personal Interviewing)
Telephone interview	CATI (Computer Assisted Telephone Interviewing)
Self-administered form:	
With interviewer present	CASI (Computer Assisted Self Interviewing); CASI (computer assisted self interviewing). CASI-V (question text on screen only: visual). CASI-A (text on screen and also heard on audio)
Mail survey equivalent	DBM (Disk by Mail) and EMS (Electronic Mail Survey), Web survey, Internet survey
Panel research	CAPAR (Computer Assisted Panel Research), Internet-panel
Various forms (no interviewer)	TDE (Touchtone Data Entry), VR (Voice Recognition), or ASR (Automatic Speech Recognition)

For the traditional mail survey, computer assisted equivalents have been developed too. Disk-by-Mail has been now used on a regular basis, and methodological knowledge on how to implement a successful Disk-by-Mail survey is available (see for examples, Ramos, Sedivi, & Sweet, 1998; Van Hattum & De Leeuw, 1999). In a Disk-by-Mail survey (DBM) a disk containing the questionnaire and a self-starting interview program is mailed to the respondent via the postal service. The respondent runs the program on his or her own computer and returns the diskette containing the completed questionnaire. Disk-by-mail is now being more and more replaced by electronic mail and web surveys. Electronic mail surveys (EMS) or Internet (web)surveys differ from DBM in the sense that respondents receive the request and return the survey data electronically. This is a field still very much in development. At present

EMS is only possible with special populations because not everyone has access to Internet. But the experience so far is positive and a multimode approach has proved to be successful in overcoming noncoverage problems caused by the limited access to Internet for subgroups (Couper, 2000, Dillman, 2000). Another way to overcome the limited computer access in web surveys, is computer assisted panel research. A panel of households is selected and the research institute provides the panel members with computers and communication equipment. Surveys are sent electronically to the household members on a regular basis, and after completion are sent back automatically. This approach proved successful for consumer panels in the Netherlands (Saris, 1998).

One of the main reasons that computer assisted data collection has become popular so quickly was the general expectation that it would improve data quality and efficiency and reduce costs. In the last two decades these claims have been investigated mainly through empirical mode comparisons of computerized and paper-and-pen versions of the same questionnaire. These studies mainly focus on data quality, only few studies also investigate costs. In the remainder of this chapter I will concentrate on data quality.

I start with a theoretical model for the influence of computer assisted interviewing, discriminating between technological and methodological data quality. I will proceed with a short overview of empirical evidence for technological data quality, timeliness and cost reduction. I will then focus on methodological data quality: what happens in the interview situation and how it influences data quality. Since acceptance of computer-assisted methods is an important criterion by itself, I also include research on the attitudes and opinions of interviewers and respondents. I end with a discussion on the challenges that new emerging technologies offer.

4.2. Survey Data Quality and Computer-Assisted Interviewing

As early as 1972, Nelson, Peyton, and Bortner pointed out that automatic routing to the next question and range checks on the given answers would enhance data quality. They emphasize *technological* or *operational* data quality: the reduction of previously required post interview data processing activities (Nicholls, 1996). Operational data quality is affected by all the technological possibilities of computer assisted interviewing.

Factors associated with the visible presence of a computer and its effect on the interview situation may affect data quality, apart from the technical aspects. These factors affect *methodological* data quality, defined by an absence of nonsampling survey bias and error.

Recently, Total Quality Management (TQM) has received much attention in industrial settings and in a lesser degree in statistical and survey establishments (see also chapter 2). As a consequence, additional criteria for 'good' data collection methods have been formulated. The most important are *timeliness and costs*. In other words, does a new technology provide the data more quickly than the old one and does it reduce the survey costs?

4.2.1. Potential for Improving Technological Data Quality

Compared to an optimally implemented paper-and-pencil interview, the optimally implemented computer assisted interview has five apparent advantages.

(1) There are no routing errors. If a computer system is correctly programmed, routing errors, that is, errors in the question order, skipping and branching, do not occur. Based on previously given answers the program decides what the next question must be, and so both interviewer and respondent are guided through the questionnaire. Missing data because of routing and skipping errors does not occur. Also, questions that do not apply to a specific respondent are automatically skipped. As a result, automatic routing reduces the number of data errors.

(2) Data can be checked immediately. An optimally implemented interview program will perform some internal validity checks. The simplest checks are range checks that compare the given response to the range of possible responses. Thus the program will refuse the response '8' if there are only seven possible response categories, and then ask to correct the response. Range checks are straightforward when the question has only a limited number of response categories. More complicated checks analyze the internal consistency of several responses. Consistency checks are more difficult to implement; one must anticipate all valid responses to questions, list possible inconsistencies, and devise a strategy for the program to cope with them. In a paper-and-pencil study, internal validity checks have to be conducted at the data cleaning stage that usually follows the data collection stage. However, when errors are detected, they can only be recoded to a missing data code because it is no longer possible to ask the respondents what they really meant. During a computer-assisted interview there is an opportunity to rephrase the question and correct range and consistency errors. This should lead to fewer data entry errors and missing data.

(3) The computer offers new possibilities to formulate questions. One example is the possibility to randomize the order of questions in a scale, giving each respondent a unique question order. This will eliminate systematic question order effects. Response categories can also be randomized, which avoids question format effects. The computer can also assist in the interactive field coding of open questions using elaborate coding schemes, which would be unmanageable without a computer. Finally, the computer can be used to employ question formats such as drawing line lengths as in psychophysical scaling, which in paper and pencil methods are more awkward to use.

(4) There is no separate data entry phase. This means that no extra errors are added. It also implies that the first tabled results can be available soon after the data collection phase. On the other hand, construction, programming, and checking of the questionnaire will take considerable time in computer-assisted data collection. Thus, a well-planned computer-assisted survey has a real advantage when the results must be quickly available right after data collection (as in election forecasts).

4.2.2. Potential for Improving Methodological Data Quality

The visible presence of a computer may affect data quality, apart from the technical aspects of using a computer. Both usability (for instance, human-computer interface, learning) and psychological factors can play a role. As with most technological innovations the effects are for the most part temporary. After some time, one gets used to the new machine, and its influence on the situation becomes smaller. Now we are clearly in a transition period; the computer is no longer an unimaginable technological wonder, but it is also not yet a common household item for everyone. Even in the USA, one of the early adapters to technological innovations, computer access is still limited but growing. In 1998, the US Bureau of the Census estimated that 26% of the households used Internet at home. The percentage of persons using internet (inside and outside the home) was estimated at 33% or a third of the adult population of 18 years and older, which corresponds to 65 million US adults (Couper, 2000)

Compared to traditional paper and pencil methods, the presence of a computer could lead to the following effects (positive and negative) on how the whole data collection procedure is perceived.

(1) Less privacy. When one is totally unfamiliar with computers there could be a 'big brother' effect, leading to more refusals and socially desirable answers to sensitive questions. When researchers first started to use computer assisted data collection, this was a much-feared effect.

(2) More privacy. Using a computer could also lead to the expectancy of greater privacy by the respondents; responses are typed directly into the computer and cannot be read by anyone who happens to find the questionnaire. Much depends here on the total interview situation and how the survey is implemented.

(3) Trained interviewers may feel more self-confident using a computer, and behave more professionally. This in turn could lead to more confidence and trust of the respondent in the interviewing procedure as a whole.

(4) The knowledge that the system accurately records information about the interview process itself (e.g., time and duration of the interview, the interval between interviews and the order in which they are carried out) inhibits interviewers to 'cheat'.

(5) The use of a computer may distract interviewers. They have to pay attention to using the computer correctly and typing in the answers accurately. If interviewers cannot touch-type, typing in long answers may lead to less eye contact between interviewers and respondents, causing the interviewers to miss nonverbal reactions of the respondents. If the computer is located between the interviewer and the respondent, even the physical distance may be greater than in a paper and pen interview. These factors all weaken the ‘rapport’ between interviewer and respondent; as a consequence the interview may not be conducted optimally, and data quality may suffer.

(6) On the other hand, a well-trained and experienced interviewer can rely on the computer for routing and complex question sequences, and therefore pay more attention to the respondent and the social processes involved in interviewing.

4.2.3. Potential for Increased Timeliness and Reduced Costs

Going from paper-and-pencil to computer assisted interviewing asks for initial investment, not only in equipment, but also in time. One has not only has to invest in hardware, and software, but also in acquiring hardware- and software-related knowledge and skills or even new personnel. In addition, basic interviewer training now needs to include training in handling a computer and using the interviewing software.

After the initial investments are made, a computer-assisted survey may be less costly and quicker than traditional data collection, but it all depends on the study: its complexity, its size, and its questionnaire. To evaluate the cost efficiency and timeliness of a computer assisted survey, a distinction should be made between *front-end* processing and *back-end* processing. In general, a well-designed computer assisted survey requires investing more time, effort, and money in the beginning of the research (front-end processing), time, effort, and money that is saved at the end stage (back-end processing). Especially the design and implementation of range and consistency checks costs more at the front-end, but reduces the time needed to prepare the data for the analysis (back-end) considerably. Furthermore no questionnaires have to be printed, entered, or coded.

4.3. Empirical Evidence for Improved Quality: Technological Data Quality, Timeliness and Cost.

4.3.1. Technological Data Quality

Technological data quality was defined above as the reduction of previously required post interview data processing activities. Using a well-programmed and tested interview program can reduce the number of errors in the data by preventing mistakes. Empirical studies confirm this expectation.

Computer Assisted Telephone Interviewing (CATI).

In their review of telephone surveys, Groves and Nicholls (1986) conclude that CATI leads to less missing data because it prevents routing errors. For instance, far more skip errors are found in paper telephone surveys then in CATI. It is not therefore not surprising, that post hoc data cleaning finds more errors with traditional paper-and-pencil methods than with CATI. However, no difference is found in respondent induced missing data (that is, 'do-not-know' and 'no-answer' responses). The same conclusions were confirmed in later studies. For an overview see Nicholls, Baker, and Martin (1997).

There are no differences in closed questions, but how about open questions? Catlin and Ingram (1988) studied the possible effects of computer use on open questions; they found no differences between computer assisted and paper interviewing. CATI performed as well as paper and pencil interviewing both in codability of the answers to open questions and in length of answers given to open questions.

Computer Assisted Personal Interviewing (CAPI)

The percentage of missing data is clearly lower in CAPI as in CATI, mostly because interviewers cannot make routing errors. There is *no* evidence that CAPI also diminishes the number of missing data caused by respondents, such as explicit 'do-not-know', or 'no-answer' (Baker & Bradburn, 1992; Martin, O'Muirchetaigh, & Curtice, 1994).

Little is known about data quality regarding open questions. Baker (1992) summarizes a study by the French National Institute for Statistical and Economical Research (INSEE) that did not find any difference between PAPI and CAPI in this respect.

Computer Assisted Self Interviewing (CASI, CSAQ)

Computer Assisted Self Administered Questionnaires (CSAQ) and Computer Assisted Self Interviewing (CASI) make it possible to use very complex questionnaires without the aid of an interviewer. But also in standard, less complex self-administered questionnaires, CASI reduces item nonresponse (see Ramos, et al, 1998). For example, in a well-designed and thoroughly tested computer questionnaire, it is impossible for a respondent to skip a question by mistake. Less mistakes lead to less item nonresponse. This is clearly illustrated by the findings of Van Hattum & De Leeuw (1999). They used computer assisted self administered questionnaires in primary schools and compared data from paper and pencil (PAPI) self administered questionnaires with data from computer assisted self administered questionnaires (CSAQ). In the CSAQ-condition the mean percentage of missing values was 5.7% (standard deviation= 3.4%), while in the PAPI-condition the mean of the percentage missing values was 14.1% (standard deviation= 25.0%). It is interesting to note that not only the average amount of missing data is less in computer assisted data collection, but also that the individual differences, indicated by the standard deviation, are smaller. Van Hattum & De Leeuw (1999) attribute this to the fact that with a paper questionnaire children who are not concentrating on the task or who are careless can easily skip a question or even a whole page by mistake, while CSAQ forces children to be more precise.

A small number of studies have explicitly compared respondent entry errors in computerized versus paper and pen questionnaires. Fewer respondent errors are reported in CASI than in paper and pen self-administered questionnaires. For an overview, see Nicholls et al (1997).

4.3.2. Timeliness and Costs

When comparing timeliness and costs, a distinction should be made between front-end processing and back-end processing. In general, front-end processing (that is, developing, implementing and testing the questionnaire) takes more time and is therefore more expensive. On the other hand, no data-entry is needed and data editing and data cleaning take less time: back-end processing is faster. With very large surveys this will save time. In general, there is no difference in the total time needed for the research. But once the interviewing has started, results are available much faster than in traditional paper-and-pencil interviewing. When timeliness and a fast release of results are important for a client, this is an important advantage of computer-assisted data collection over paper-and-pencil methods. During interviewing, time may be saved by the improved efficiency of computer assisted sample management (Nicholls & De Leeuw, 1996).

Computer Assisted Telephone Interviewing (CATI).

Most studies that attempt to weigh the costs and advantages of CATI conclude that the initial investments in hardware and software pay off only for large scale or regularly repeated surveys. A rule of thumb is that the break-even point is at about thousand telephone interviews. Below that number, the argument of cost reduction is, by itself not sufficient to use CATI (Weeks, 1992).

Computer Assisted Personal Interviewing (CAPI)

CAPI requires a larger investment in hardware, software and support staff than CATI. These high fixed costs are only compensated by lower flexible costs. For instance, the cost savings in printing, data-entry, and editing for large scale surveys. There is limited empirical data on cost comparisons between computer assisted and paper and pencil personal interviews. Two studies systematically assess costs for CAPI: initial investment in hardware and software was excluded, but extra fieldwork costs for training and supervision were included. Sebestik et al. (1988) compared costs in a small scale CAPI experiment in the USA. Their conclusion is that overall CAPI was more expensive, mostly because of added costs in training and supervising interviewers. In a larger experiment Baker and Bradburn (1992) conclude that CAPI was still more expensive (±12%) than PAPI; the cost reduction in entering and cleaning data was not large enough to offset the higher training and supervision costs. Baker extrapolates these findings and concludes that when fixed hardware costs are *excluded*, approximately 1500 CAPI interviews are needed to reach the break-even point between increased front-end and decreased back-end costs.

Computer Assisted Self Interviewing (CASI, CSAQ)

Computer assisted self-administered questionnaires (CSAQ) and Disk-by-Mail and e-mail surveys have the advantage that no interviewers are needed, so in comparison with CATI and CAPI they save costs. This is one of the main reasons why Baker (1998) predicts a decline of interviewing and a rise of CASI and CSAQ. When one compares computer assisted procedures with the traditional paper mail survey cost savings are not so obvious. As with all forms of computer assisted data collection, the extra investment in programming the questionnaire and debugging only pays off for large surveys where printing and editing make the paper form more costly (see also. Ramos, et al, 1998). In Disk-by-Mail, the mailing costs include a special protective envelope. Also, a disk is heavier than a short paper questionnaire, which makes DBM in generally somewhat more costly than paper mail questionnaires. However, when large numbers of longer questionnaires have to be mailed, DBM can be a real cost saver. Van Hattum and De Leeuw (1999) systematically compare costs for a DBM and a paper mail survey of 6000 pupils in primary schools. They conclude that the average cost for a completed questionnaire is 1.01 US dollars for a Disk-by-Mail survey and 3.22 US dollars for a paper-and-pen mail survey.

E-mail and web surveys pose an extra challenge for Europe. Clayton and Werking (1998) describe the cost savings (for instance, labour, postage) in a e-mail survey of businesses. Transmission costs (telephone) are practically zero. However, unlike the USA, in most European countries local telephone calls are not free! This not only increases the costs for the researcher, but also increases the costs (internet connect time both receiving and sending) for the potential respondent. To ensure high response rates, one should find ways to reduce respondent costs comparable to prepaid return postage in mail surveys, or reimburse factual costs.

4.4. Empirical Evidence for Improved Quality: Acceptance of New Technology and Methodological Data Quality.

4.4.1. Acceptance of New Technology

The use of a computer may have an influence on the behaviour of both interviewer and respondent. Therefore, in the first applications of computer assisted interviewing special attention was paid to the acceptance of the new technology.

Acceptance by interviewers CATI and CAPI

In the early days, when systems were slow and portable computers heavy, interviewer acceptance was not general. Acceptance depended strongly on the speed and reliability of systems (Nicholls et al, 1997; De Leeuw et al, 1998). With modern systems acceptance is high. Well-trained interviewers are positive about computer-assisted interviewing. They appreciate the support that a good system offers when complex questionnaires are employed, they like working with the computer, and derive a feeling of professionalism from it. However, crucial for acceptance is that interviewers are well-trained in general computer skills, in the specific computer assisted interview system that is used, and in general interviewing techniques. For training interviewers in computer assisted interviewing, see Woijcik & Hunt (1998). Besides training, ergonomic factors are of influence too: readability of screens, well-defined function keys, and usability, are important factors for acceptance.. In addition, a good human-computer interface may contribute to the avoidance of human errors.

Acceptance by respondents and unit nonresponse: CATI and CAPI

In telephone interviews, respondents as a rule will not notice whether a computer is used or not, it is not surprising that no differences in unit nonresponse are found between CATI and traditional paper and pen telephone interviews. When computer assisted personal interviewing was introduced researchers were afraid of a negative effect on response rates. But even in the first applications of the method in Sweden and the Netherlands this did not occur. Later studies confirm that CAPI and paper-and-pencil methods yield comparable response rates in studies in the U.S.A., England, Sweden , and Germany. For an overview, see De Leeuw (2002). These studies also report very low percentages of spontaneous negative reactions by respondents (1-4%). Most reactions are neutral or positive.

When respondents are explicitly asked for a reaction to using the computer they generally react positively and are found to prefer the computer-assisted form: most respondents find CAPI interesting, and attribute a greater degree of professionalism to CAPI. The social interaction with the interviewer is generally described as comfortable and relaxed. Only a small percentage (5%) reports negative feelings (see also, De Leeuw, 2002).

Acceptance by respondents and unit nonresponse: CASI

Various forms of computer assisted self-administered questionnaires appear to be appreciated by the respondents; they evaluate it positively and find it interesting and easy to use (for overviews see Ramos et al, 1998; De Leeuw et al, 1998). Beckenbach (1995) reports that more than 80% of

the respondents had no problem at all using the computer or the interviewing program, and that few respondents complained about physical problems such as eye-strain.

The general positive appreciation also shows in the relative high response ratio with Disk By Mail (DBM) surveys at the end of the twentieth century and it was not unusual to have response ratio's of 40 to 50 percent without using any reminders. Assuming that this is a special population interested in the research topic, an ordinary well conducted mail survey using no reminders may be expected to yield about 35% response (see also, De Leeuw, 2002). The high response rates may be partly caused by the novelty value of DBM at that time, and will diminish over time. It should be noted that Ramos et al (1998) found no evidence for higher response rates in DBM in academic and government surveys.

How e-mail or web surveys will develop remains unsure. The novelty value is wearing of, and electronic junk-mail is increasing. Also, one mouse-click is enough to through away anything unwanted or uninteresting. In addition, there are added financial costs (connect time) for respondents in most European countries. This could lead to extreme low response rates, which would threaten the validity of the conclusions. To ensure an acceptable response for e-mail and web surveys one should carefully analyze what makes electronic surveys different (for instance, regarding security of the net, costs). These issues should be carefully addressed, in doing this we can learn from the past. Many principles that in the past have proved to be successful in paper mail surveys, can be successfully translated to electronic surveys (Dillman, 2000). But, we have to go one step further, we must learn to optimally use the enormous audio-visual potential of this new medium (Couper, 2000), and learn to design questionnaires while using visual language (Dillman, 2000).

At present, there are promising results from panel-surveys, which use internet. In the Netherlands at Tilburg University a *general population household* panel is now completely operating through internet. Of course, panel members received instruction in how to use the new technology, a help-desk is available, and all costs are reimbursed (Sikkel, 2002)

4.4.2. Methodological Data Quality

Computer Assisted Telephone Interviews (CATI)

In telephone interviews the computer is not visible present. Respondents may occasionally hear keyboard clicks, or be told by the interviewers that a computer is used. No systematic research has been done on the effects of this knowledge, but the general impression is that it makes no difference to respondents if they know that their answers are typed directly into a computer. It is therefore not surprising that there are no indications for any differences in methodological data quality between computer assisted and paper and pen telephone interviews. CATI does lead to less missing data because it prevents routing errors, but there is **no** difference in *respondent induced* missing data because of 'don't know' and 'no answer' responses. Also, no differences in 'openness' or social desirability are found (see: Groves & Nicholls, 1986; Weeks, 1992).

Interviewers, however, know that a computer system is used, and that more rigid control takes place. Computer assisted interviewing often leads to a greater standardization of the interview, to the extent that interviewers sometimes complain about 'rigidity'. In general, researchers appreciate this greater standardization because this minimizes interviewer bias, and the greater standardization can be seen as an advantage of CATI too. There is indeed some confirmation of

more standardization of interviewer behaviour in CATI: in a controlled comparative study, using the same interviewers both for traditional and for computer assisted interviews, Groves and Mathiowetz (1984) found less interviewer variance in CATI than in the paper-and-pencil telephone interviews.

Computer Assisted Personal Interviewing (CAPI)

In face-to-face interviews the computer is highly visible and respondents may react to its presence. This could influence respondents' trust in the privacy of the data. When researchers first started to use CAPI, they feared a 'big brother' effect, leading to more refusals and socially desirable answers to sensitive questions. An alternative hypothesis was that the use of a computer could also lead to feelings of greater privacy by the respondents; responses are typed directly into the computer and cannot be read by anyone who happens to find the questionnaire. There is no hard empirical evidence for either hypothesis. The acceptance of computer assisted face-to-face interviewing is high for both respondents and interviewers, and there are no indications that using a computer disturbs the interviewing situation (Beckenbach, 1992; see also De Leeuw, 2002).

An early and much cited comparative study by Waterton (cf. Waterton & Duffy, 1984) reports a positive effect of CAPI with a sensitive question about alcohol consumption; using the CAPI method more alcohol consumption was reported, which means that presumably CAPI was less affected by social desirability bias. However, in the CAPI mode the sensitive question was asked by letting the respondent type their own answers into the computer, unseen by the interviewers, which makes this part of the interview like a self-administered questionnaire (CASI). In the traditional paper and pen mode, the question was asked by the interviewer and the answer was taken down by the interviewer. Since self-administered questionnaires typically show less social desirability bias than face-to-face interviews (for an overview, see De Leeuw, 1992), the reported difference between PAPI and CAPI in this study may well correspond to a difference between an interview and a self-administered questionnaire, and not to a technology effect.

Studies that do compare paper and pen face-to-face interviewing and computer assisted personal interviewing and therefore focus on the effect of the new technology more purely, do report slightly less social desirability bias with CAPI (see for instance, Baker & Bradburn, 1992; Martin et al., 1994; Tourangeau & Smith, 1998). However, the differences are very small; furthermore, Tourangeau and Smith (1998) report an interesting interaction with location of interview. When the interview took place in the respondent's home, the computer assisted version produced more 'openness' in answers. However, when in a health clinic, fewer open answers were given and the computer assisted version revealed fewer sex partners than the paper and pen version. This suggests that setting is important. It is more the way respondents perceive the total (computer assisted) interview situation, than the use of the computer itself, that influences methodological data quality (De Leeuw, 2002).

Computer Assisted Self Interviewing (CASI, CSAQ)

There is strong evidence that for paper-and-pen modes, self-administered questionnaires are better at eliciting sensitive information than interviews (for an overview, see De Leeuw, 1992; De Leeuw & Collins, 1997). Computer-assisted self-interviewing has the additional advantage that complex questionnaires with many routings (e.g., health inventories) can now be administered in self-administered form. Whether a computer-assisted form also will produce

more open answers and more self-disclosure than a paper and pen questionnaire has been the topic of a number of studies.

Several studies showed more self-disclosure on sensitive topics (such as, abortion, male-male sexual contact) when using CASI. There is some evidence that the use of Audio-CASI does not chance this effect (for an overview, see De Leeuw, 2002). In a meta-analysis of 39 studies, Weisband and Kiesler (1996) found a strong significant effect in favor of computer methods. This effect was stronger for comparisons between CASI and face-to-face interviews. But, even when CASI was compared with self-administered paper-and-pen questionnaires, self-disclosure was significantly higher in the computer condition. The effect reported was larger when more sensitive information was asked. Weisband and Kiesler (1996) also report the interesting finding that the effect is diminishing over the years, although it has not disappeared! They attribute the diminishing effect to a growing familiarity with computers and their possibilities among the general public. Interestingly, their meta-analysis also showed that the data were not homogeneous. This means that although the general trend was in favor of computer assisted methods, some studies showed the opposite effect.

Recent research suggests that these contradictory findings could be attributed to the interview situation and perceived privacy (see De Leeuw, 2002). Therefore, when using computer assisted questionnaires one should take careful precautions to gain respondents' trust. The setting and the implementation of the questionnaire should reassure the respondent about confidentiality. Simple precautions, like masking the answer or refreshing the screen when the answer has been given, will probably do the trick. Also, whenever, other persons are in the same room - be it interviewers, family members, teachers, or other students in a lab - they should be kept at some distance.

Hardly any empirical tests have been made regarding electronic mail surveys or web surveys. A recent study in Japan, emphasizes the importance of mutual trust for (non)response and data quality (Yoshimura & Ohsumi, 2000). In an early study in the USA, Kiesler and Sproull (1986) found fewer socially desirable answers in an electronic questionnaire than in the paper mail version. Subsequent studies (Mitchell, 1993) found no differences. Extrapolating the findings summarized above on both CASI and internet surveys, I suggest that in e-mail and web surveys *privacy and security* could be crucial factors when asking for sensitive information. Respondents should have the feeling that their answers are safe, and encryption in combination with an icon to convey the message should be standard. When designing special surveys, we should focus more on the human-computer interaction and the perceptions and reactions of the respondent. In the end it is the respondent not the technology that matters.

4.5. Summary

Computer assisted telephone interviewing, and to a lesser degree, computer assisted face-to-face interviewing, are by now widely used in survey research. Computer-assisted self-interviewing is gaining in popularity and other forms of electronic data collection, such as web-surveys, and electronic data exchange, are now used and studied with enthusiasm.

Computer assisted data collection has a high potential to improve data quality. This together with the expectations that it would also improve efficiency and reduce costs, was why computer assisted data collection has become popular so quickly. However, for most of these potential advantages the empirical evidence is still limited. Systematic comparisons of *costs*

and *efficiency* are rare, and the evidence for cost and time reduction is not very strong. A well designed-computer assisted survey requires investing more time, money, and effort in the beginning of the process (front-end processing), which is saved at the end stage (back-end processing. These investments will only pay off in large scale or regularly repeated surveys.

There is little evidence that the use of CAPI, CATI and Disk-by-Mail surveys improves *response rates*. Conversely, there is also no evidence for a decrease in response rates. How e-mail and web surveys will develop remains uncertain. The novelty value is wearing off and electronic junk-mail is increasing. To ensure an acceptable response and good data quality, one should carefully analyse what makes web surveys different (for example, security, access to the net of different demographic groups, connect costs, influence of screen lay-out on measurement error and the robustness of lay-out for different web-browsers), and address these issues in the design of the survey.

There is ample empirical research of improved *technological data quality* in computer-assisted data collection. A well-programmed and tested interview program will have range and consistency checks, and prevent routing errors, which results in far less item nonresponse. Computer assisted data collection is no panacea for good data quality. It requires one to do almost everything that is needed with a good paper-and-pen interview or questionnaire, and to add extra effort in computer implementation, in testing the questionnaire, in visual design and designing an ergonomic screen lay-out, in extra interviewer training, and in designing a respondent friendly and trustworthy questionnaire. However, this investment is earned back in far less interviewer error and the error-free administration of complex questionnaires.

There is some evidence that computerized methods of data collection improve *methodological data quality*. Respondents are less inhibited and show more self-disclosure when sensitive questions are used. But this effect may be diminishing over time, as some studies suggest. Furthermore, there is evidence that much depends on the perception of the interview situation by the respondent and on careful design of the total study and of the computer interface. For instance, the distance between computers in a computer lab influences the openness on answers; a larger distance gives more openness. Also, whether or not the typed-in answers remain on the screen or are 'masked', and whether sounds come over a headphone or through speakers in computer assisted self-interviews may affect answers on sensitive questions.

Systematic research on these topics will teach us more about how to use computers optimally in data collection. In doing this, we should keep in mind that it is the human that counts not the technology. How respondents perceive the interview situation, how large their (mis)trust in computers is and how much they trust the survey organization, will determine the success of computer-assisted method and especially of web surveys.

Finally, I should emphasize that computer assisted data collection is no panacea for good data quality. It requires one to do almost everything that is needed with a good paper-and-pen interview or questionnaire, and to add extra effort in computer implementation, in testing the questionnaire, in designing an ergonomic screen lay-out, in extra interviewer training, and in designing a respondent friendly and trustworthy questionnaire. However, this investment is earned back in far less interviewer error and the error-free administration of complex questionnaires. It also offers us the opportunity to use questionnaires with complex routing patterns, without the help of an interviewer.

4.6. Suggested Readings and Websites

Handbooks:

Don A. Dillman (2000). Mail and Internet Surveys; The Tailored Method. New York: Wiley

Mick Couper et al (1998). Computer Assisted Survey Information Collection. New York: Wiley.

Web sites:

On web surveys:

The web page of the web survey methodology project at http://www.websm.org/

On general survey methods:

Webpage Australian Statistical Office at http://www.sch.abs.gov.au/ Especially the section research papers

References

Baker, R.P. (1992). New technology in survey research: Computer assisted personal interviewing (CAPI). *Social Science Computer Review*, 10, 145-157.

Baker, R.P. (1998). The CASIC future. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection* (pp. 583-604). New York: Wiley.

Baker, R.P. & Bradburn, N.M. (1992). CAPI: Impacts on data quality and survey costs. *Information Technology in Survey Research Discussion Paper 10* (also presented at the 1991 Public Health Conference on Records and Statistics).

Beckenbach, A. (1992). *Befragung mit dem Computer, Methode der Zukunft?. Anwendungsmöglichkeiten, Perspektiven und experimentelle Untersuchungen zum Einsatz des Computers bei Selbstbefragung und persönlich-mündlichen Interviews.* [In German: Computer Assisted Interviewing, A method of the future? An experimental study of the use of a computer by self-administered questionnaires and face to face interviews]. Ph.D. thesis. Universität Mannheim.

Bretschneider, M., & Schumacher, J. (1996). DEMOS Eine Datenbank zum Nachweis Kommunaler Umfragen auf dem Weg zum Analyse instrument. *ZA-Information, 38*, 59-81.

Biemer, P.P., and Lyberg, L.E. (2003). *Introduction to survey quality.* New York: Wiley.

Catlin, G. & Ingram, S. (1988). The effects of CATI on costs and data quality: A comparison of CATI and paper methods in centralized interviewing. In: R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II & J. Waksberg (Eds.). *Telephone survey methodology.* New York: Wiley.

Clayton, R.L. & Werking, G.S. (1998). Business surveys of the future: The world wide web as a data collection methodology. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection* . New York: Wiley.

Collins, M., Sykes, W., O'Muircheartaigh, C. (1998). Diffusion of technological innovation: Computer assisted data collection in the UK. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection* . New York: Wiley.

Couper, M.P., & Nicholls, W.L.II (1998). The history and development of computer assisted survey information collection methods. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection* . New York: Wiley.

Couper, M.P. (2000). Websurveys; A review of issues and approaches. *Public Opinion Quarterly, 64*, 4, 464-494. See also Couper (2000) the Good, the Bad, and the Ugly. University of Michigan, Institute for Social Research, Survey Methodology Program, Working paper series # 077.

Couper, M.P. & De Leeuw, E.D. (2003) Nonresponse in cross-cultural and cross-national surveys. In: J.A. Harkness, F.J.R. van de Vijver, & P.P. Mohler (eds). *Cross-cultural survey methods.* New York: Wiley.

R. Czaja & J. Blair (1996). *Designing surveys: A guide to decisions and procedures.* Thousand Oaks: Sage (Pine Forge Press series in research methods and statistics).

R. Czaja & J. Blair (2005). *Designing surveys: A guide to decisions and procedures, second edition.* Thousand Oaks: Sage (Pine Forge Press series in research methods and statistics).

Deming, W.E. (1944), On Errors in Surveys, *American Sociological Review, 9*, 359-369.

Deming,W.E. (1982). *Quality, productivity and competitive position*. Cambridge, Ma: Massuchusetts Institute of Technology.

Dillman, D.A. (1978). *Mail and telephone surveys; The Total Design Method*. New York: Wiley.

Dillman, D.A. (2000). *Mail and internet surveys; The Tailored Design Method*. New York: Wiley.

Dillman, D.A., Eltinghe, J.L., Groves, R.M., & Little, R.J.A. (2002). Survey nonresponse in design, data collection, and analysis. In: R.M. Groves, D.A. Dillman, J.L. Eltinge, R.J.A. Little (eds). *Survey nonresponse.* New York: Wiley

Dillman, D.A., & Christian, L.M. (2003). Survey mode as a source of instability in responses across surveys. *Field Methods, 15*, 2, 1-22.

Eurostat (2000).*Assessment of the quality in statistics*. General standard report, Luxembourg, April 2000 at http://www.unece.org/stats/documents/2000/11/metis/crp.3.e.pdf (accessed January 2005)

Fowler, F.J. Jr. (1991). Reducing interviewer-related error through interviewer training, supervision, and other means. In: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S.Sudman (Eds). *Measurement errors in surveys*. New York: Wiley.

Fowler, F.J. Jr. (1995). *Improving survey questions: Design and evaluation*. Thousands Oaks: Sage Applied Social Research Methods Series Vol 38

Gaziano, C. Comparative analysis of within-household respondent selection. *Public Opinion Quarterly, 69,* 1, 124-157.

Goyder , J. (1987). *The silent minority; Nonrespondents on sample surveys*. Cambridge: Policy Press.

Groves, R.M. (1989). *Survey errors and survey costs.* New York. Wiley.

Groves, R.M. (1991). Measurement error across disciplines. In: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S.Sudman (Eds). *Measurement errors in surveys*. New York: Wiley.

Groves, R.M. & Mathiowetz, N.A. (1984). Computer assisted telephone interviewing: Effects on interviewers and respondents. *Public Opinion Quarterly, 48*, 356-369.

Groves, R.M. & Nicholls, W.L. II (1986). The status of computer-assisted telephone interviewing: Part II-Data quality issues. *Journal of Official Statistics, 2*, 117-134.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953), *Sample Survey Methods, Vol. II,* New York: Wiley.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1961), Measurement Error in Censuses and Surveys, *Bulletin of the International Statistical Institute, 38*, 359-374.

Van Hattum, M., & De Leeuw, E.D. (1999). A Disk-by-Mail survey of teachers and pupils in Dutch primary schools; Logistics and data quality. *Journal of Official Statistics (JOS), 15,* 3, 413-429.

Heberlein, T.A., & Baumgartner, R.M. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review, 43*, 447 462.

Hox, J.J. (1986). *Het gebruik van hulptheorien bij operationaliseren* [In Dutch: Using auxiliary theories for Operationalization: A study of the construct of subjective well-being]. Amsterdam: University of Amsterdam

Hox, J.J, & De Leeuw, E.D. (1994).A comparison of nonresponse in mail, telephone, and face-to-face surveys. *Quality and Quantity, 28*, 329-344.

Hox, J.J. (2002). *Multivel analysis; techniques and applications.* Mahwah: Erlbaum

Kiesler, S. & Sproull, L.S. (1986). Response effects in electronic surveys. *Public Opinion Quarterly, 50,* 402-413.

Kish, L. (1965), *Survey Sampling,* New York: Wiley.

Kruskal, W. (1991). Introduction to measurement errors in surveys. In: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S.Sudman (Eds). *Measurement errors in surveys*. New York: Wiley.

De Leeuw, E.D., & Van der Zouwen, J. (1988).Data quality in telephone and face-to-face surveys: A comparative meta-analysis, In : R. Groves, P. Biemer, L. Lyberg, J. Massey., W. Nicholls II, & J. Waksberg (eds). *Telephone survey methodology*. New York: Wiley.

De Leeuw, E.D. (1992). *Data quality in mail, telephone and face to face surveys*. Amsterdam: TT-Publikaties.

De Leeuw, E., & Nicholls, W.L. II (1996). Technological innovations in data collection: Acceptance, data quality and cost. *Sociological Research Online, Vol 1*, no 4, <http:/www.socresonline.org.uk/socresonline/1/4/leeuw.html>

De Leeuw, E.D. & Collins, M. (1997). Data collection method and data quality: An overview. L. Lyberg, P. Biemer, M. Collins, C. Dippo, N. Schwarz, & D. Trewin (eds). *Survey Measurement and Process Quality*. New York: Wiley.

De Leeuw, E.D., Hox, J.J., & Snijkers, G. (1998). The effect of computer-assisted interviewing on data quality. In: B. Blyth (ed). *Market research and information technology; Application and innovation.* Esomar monograph 6. Amsterdam: Esomar.

De Leeuw, E.D. (2002). The effect of computer assisted interviewing on data quality: A review of the evidence [CD-Rom]. In: J. Blasius, J. Hox, E. de Leeuw & P. Schmidt (Eds.), *Social science methodology in the new millennium*. Opladen, FRG: Leske + Budrich

De Leeuw, E.D., & De Heer, W. (2001). Trends in household survey nonresponse: A longitudinal and international comparison. In: R.M. Groves, D.A. Dillman, J.L. Eltinge, R.J.A. Little (eds). *Survey nonresponse*. New York: Wiley

De Leeuw, E.D., Hox, J., Huisman, M. (2003). Prevention and treatment of item nonresponse. *JOS, (Journal of Official Statistics) 19*, 153-176.

Lyberg et al (1991) *Summary report from the Leadership Group (LEG) on Quality* at http://amrads.jrc.cec.eu.int/WPs%20pages/Quality/Documents/LEGsummary.pdf (accessed January 2005)

Makkai, T. & McAllister, I. (1992). Measuring social indicators in opinion research; A method to improve accuracy on sensitive questions. *Social Indicators Research, 27*, 2, 169-186.

Martin, J., O'Muircheartaigh, C. & Curtice, J. (1994). The use of CAPI for attitude surveys: An experimental comparison with traditional methods. *JOS* (*Journal of Official Statistics), 9*, 641-661.

Mitchell, D.L. (1993). *A multivariate analysis of the effects of gender and computer vs papi modes of administration on survey results.* Louisiana Tech University.

Nathan., G. (2001). Telesurvey methodologies for household surveys: A review and some thoughts for the future. *Survey Methodology, 27,* 1, 7-31

Nicholls, W.L. II, Baker, R.P., & Martin, J. (1997). The effect of new data collection technologies on survey data quality. In: L. Lyberg, P. Biemer, M. Collins, C. Dippo, N. Schwarz, & D. Trewin (eds). *Survey measurement and process quality*. New York: Wiley.

Nicholls, W.L. II, & De Leeuw, E.D. (1996). Factors in acceptance of computer assisted interviewing methods: A conceptual and historical review. Proceedings of the section of survey research methods. *American Statistical Association* (pp. 758-763).

Nicholls, W.L. II (1996). *Definition and assessment of survey data quality in measuring the effects of new data collection technologies.* WAPOR-conference, 1996; published in a abbreviated form in: Bulletin of the International Statistical Institute, proceedings of the 51st session (Istanbul), tome LVII, Book 1, 507-510.

Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E, Martin, J., Singer, E. (2004). *Methods for testing and evaluating survey questionnaires*. New York: Wiley.

Ramos, M., Sedivi, B.M., & Sweet, E.M. (1998). Computerized self-administered questionnaires. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection*. New York: Wiley.

Saris, W.E. (1998). Ten years of interviewing without interviewers: The telepanel. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection*. New York: Wiley.

Salant, P., & Dillman, D.A. (1994). *How to conduct your own survey.* 1994. New York: Wiley.

Sebestik, J., Zelon, H., DeWitt, D., O'Reilly, J.M. & McCowan, K. (1988). *Initial experiences with CAPI.* Paper presented at the U.S. Bureau of the Census Annual Research Conference, Washington, D.C.

Schwarz, N., Strack, F., Hippler, H-J, & Bishop G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5,* 193-212.

Sikkel, D. (2002). Computer Assisted Interviewing in Longitudinal Surveys. In: J. Blasius, J. Hox, E. de Leeuw & P. Schmidt (Eds.), *Social science methodology in the new millennium.* Opladen, FRG: Leske + Budrich

Sperry, S., Bittner, D. & Branden, L. (1991). *Computer assisted personal interviewing on the current beneficiary survey.* Paper presented at the AAPOR 1991 conference, Phoenix, Arizona.

Tourangeau, R., & Smith, T.W. (1998). Collecting sensitive information with different modes of data collection. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection* . New York: Wiley.

Vehovar, V., Batagelj, Z., Lozar Mnafreda, K., & Zaletel, M. (2002). Nonresponse in web surveys. In: In: R.M. Groves, D.A. Dillman, J.L. Eltinge, R.J.A. Little (eds). *Survey nonresponse.* New York: Wiley

Waterton, J.J. & Duffy, J.C. (1984). A comparison of computer interviewing techniques and traditional methods in the collection of self-report alcohol consumption data in a field survey. *International Statistical Review, 2*, 173-182.

Weeks, M.F. (1992). Computer-Assisted Survey Information Collection: A review of CASIC methods and their implication for survey operations. *JOS (Journal of Official Statistics), 4*, 445-466.

Weisband, S. & Kiesler, S. (1996). *Self-disclosure on computer forms: Meta-analysis and implications.* Tucson: University of Arizona .

Woijcik, M.S., & Hunt, E. (1998). Training field interviewers to use computers: Past, present, and future trends. In: M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, & J.M. O'Reilly (eds). *Computer assisted survey information collection*. New York: Wiley.

Yoshimura, O., & Ohsumi, N. (2000). *Some experimental surveys on the WWW Environments.* Presented at the session Internet surveys and retrieval, International Meeting of the International Federation of Classification Societies (IFCS), Namur, Belgium, 2000.

Erakunde autonomiaduna
Organismo Autónomo del

EUSKO JAURLARITZA
GOBIERNO VASCO

Eustat
EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADÍSTICA