Cognitive Laboratory Experiences:
On Pre-testing Computerised Questionnaires and Data Quality

ISBN 90-393-2928-1

Cover design: Jan de Looff and Ger Snijkers

References

American Statistical Association, 1999, Designing a Questionnaire. ASA series: What is a Survey? (ASA, Alexandria, VA.)

Clark, H., and M. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), 1992, Questions about Questions: Inquiries into the Cognitive Bases of Surveys, pp. 15-48. Russel Sage Foundation, New York.)

Fowler, F.J. Jr., and A.M. Roman, 1992, A Study of Approaches to Survey Question Evaluation. (Final Report for CSMR, SRD division of the U.S. Bureau of the Census, Washington, DC 20233 U.S.A.)

Cognitive Laboratory Experiences:
On Pre-testing Computerised Questionnaires and Data Quality

Ervaringen met vragenlabonderzoek:
Over het pre-testen van gecomputeriseerde vragenlijsten en datakwaliteit

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht op gezag van
de Rector Magnificus, Prof. dr. W.H. Gispen,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op
maandag 28 januari 2002 des middags te 12.45 uur

door

Gerardus Johannes Maria Elisabeth Snijkers

geboren op 7 februari 1963, te Schaesberg

Promotoren:	Prof. dr. J.J. Hox
	Prof. dr. J.B. Billiet
Overige leden:	Prof. dr. H. ‘t Hart
	Prof. dr. P.G.M. van der Heijden
	Prof. dr. J. van der Zouwen
	Dr. H. Houtkoop-Steenstra

Dit proefschrift werd mogelijk gemaakt met steun van het Centraal Bureau voor de Statistiek.

man is the measure of all things

**“Surveyers cannot possibly write perfect questions,
self-evident to each respondent, that never need clarification.
And because they cannot, the answers will often be surprising.”**
(Clark & Schober, 1992)

**“The questionnaire designer must understand the need to
pretest, pretest, and then pretest some more.”**
(American Statistical Association, 1999)

To Doortje

Contents

Voorwoord

Op een dag in december 1991, vlak voor Kerst, riep Peter Kooiman (destijds Hoofd van de Hoofdafdeling Statistische Methoden, M1, van het Centraal Bureau voor de Statistiek) mij op zijn kamer en vroeg mij of ik ervoor voelde om, samen met Hans Akkerboom (destijds werkzaam bij M1-Heerlen) een zogenaamd vragenlab te gaan opzetten bij M1 in Heerlen. Toen Peter mij dit vroeg werkte ik al zo'n 5 jaar bij M1 te Voorburg en ik was wel toe aan een verandering.

In mei 1992 verhuisde ik naar Heerlen. Dat was het begin van een periode bij het Vragenlab die duurde tot september 2000. Een periode van 8½ jaar waarop ik terugkijk als een periode van hard werken met enthousiasme, grote inzet en vol overgave, maar ook met dieptepunten. Een periode waarin veel is veranderd in mijn leven. En ik vraag we me wel eens af, hoe mijn leven er nu uit zou zien als ik in 1991 'nee' had gezegd. Maar … ik heb toen 'ja' gezegd en daar heb ik geen spijt van!

Dit proefschrift kwam in deze periode tot stand.

In die periode heb ik veel enthousiaste mensen ontmoet, zowel binnen het CBS als daarbuiten. Mensen, die op één of andere manier een bijdrage hebben geleverd aan de totstandkoming van dit proefschrift. Een aantal van hen wil ik hier noemen en hen hiervoor danken.

Ik noemde al *Peter Kooiman*. Hij zorgde ervoor dat het Vragenlab van start ging. Hij heeft zich gedurende de genoemde periode weliswaar weinig met de ontwikkelingen bij het Vragenlab kunnen bemoeien, maar hij is zich er altijd voor blijven interesseren. Toen hij in september 2000 afscheid nam bij het CBS noemde hij het Vragenlab als één van zijn initiatieven waar hij trots op is. Peter, ik heb je interesse in mijn werk altijd gewaardeerd.

Ook noemde ik al *Hans Akkerboom*, met wie ik vele jaren bijzonder prettig heb samengewerkt. Samen zijn we aan het Vragenlab begonnen en samen hebben we het uitgebouwd. Vanaf het allereerste begin had ik het plan om op het werk bij het Vragenlab te gaan promoveren. Het proefschrift zou opgebouwd worden uit rapporten en artikelen op basis van onderzoek- en adviesprojecten. Peter en Hans steunden dit idee van harte; en zo kreeg mijn proefschrift langzaam vorm. Op de achtergrond is Hans betrokken geweest bij veel hoofdstukken van dit proefschrift. Hans, door jouw onuitputtelijke stroom van ideeën was jij steeds weer een grote stimulans voor mij.

Al vrij snel na het begin kwam *Martin Luppes* het Vragenlab versterken. Samen werkten we met veel plezier aan de eerste projecten. Helaas verliet hij het Vragenlab bij de TEMPO-reorganisatie van het CBS in september 1994, maar ook daarna hebben Martin en ik nog gezamenlijk gewerkt aan het uitdragen van onze ideeën over 'actieve berichtgeverscommuncatie',

zoals tijdens de Second International Conference on Etablishment Surveys (Buffalo, juni 2000), waar we een sessie over 'Improving Response Rates of Business Surveys using Integrated Communication Strategies' hebben georganiseerd. Aan de hoofdstukken 6 en 9 hebben we samen gewerkt. Martin, jij bent een fijne collega.

Vanaf september 1994 werd het Vragenlab ondergebracht bij de nieuw te vormen Sector Waarnemingsmethodologie, onder leiding van *Jean Ritzen*. Jean, jou wil ik bedanken voor je stimulans. Jij stimuleerde mij steeds weer bij het werken aan het -bij tijd en wijlen- traag vorderende proefschrift.

Hans and Jean also stimulated me to present papers at international conferences. There I met many foreign colleagues. Here I would like to mention *Pamela Campanelli, Jim Esposito* (US Bureau of Labor Statistics), *Allen Gower* (Statistics Canada), *Hakan Lindstrom* (Statistics Sweden), and *Jennifer Rothgeb* (US Bureau of the Census). To me, it has always been a great pleasure to be in your company. I met Pamela at the First International SGCSA Conference in Bristol, September 1992, where we discussed cognitive labs. I still remember the dancing party. Pam, you have been a great help (e.g. at setting-up the Questionnaire Lab and presenting a course on pre-testing methods in 2001) and a true friend during all those years. Thank you! I met Jennifer and Jim at the International Conference on Survey Measurement and Process Quality in Bristol, April 1995, where we discussed the CAQI method (chapter 5). I also do recall like yesterday how Hakan and I went to see Cambridge during the Fourth International Social Science Methodology Conference in Essex, July 1996, where I presented a paper on CAQI. With great pleasure I remember my visit to Statistics Canada and the US Census Bureau, March 1997, where I was invited by Allen and Jennifer to discuss cognitive laboratory methods. And Jennifer, Jim and I had a very pleasant discussion on gaining acceptance of cognitive laboratory research and the future of this kind of research at the Fifth International Social Science Methodology Conference in Cologne, October 2000. I still remember our diner in a Thai restaurant.

In de jaren na 1994 werden vele onderzoek- en adviesprojecten in het Vragenlab uitgevoerd. Over een aantal van deze projecten wordt in dit proefschrift gerapporteerd. De collega's waar ik in die jaren nauw mee heb samengewerkt en die als zodanig een bijdrage hebben geleverd aan mijn proefschrift, wil ik hier noemen. Dit zijn: *Dirkjan Beukenhorst, Math Bosch, Francine Dehue, Wim de Heer, Hans van Kerkoerle, Frans Kerssemakers, Isje Wolters-Kuijpers, Annemieke Luiten, Jan de Looff, Hans van Oostrum, Marko Roos* en *Leon Vankan*. Ik dank jullie allen voor de prettige en collegiale samenwerking in al die jaren. Speciaal dank ik Hans van Oostrum, bij wie ik regelmatig stoom ging afblazen, en Jan de Looff voor zijn creatieve inbreng bij het drukken van dit proefschrift.

Verder dank ik de collega's die hebben meegewerkt aan één van die projecten: *Jacques Beljon, Ad Bemelen, Piet Boersma, Wim Burgers, Piet van de Donk, Rob Giessen, Robert Göttgens, Jos Heinink, Abby Israëls, Sjon Kleintjens, Paul van Litsenburg, Branislav Mikulic, Ben Resing* en *de veld- en COTEL-interviewers van de POLS-test.*

In die projecten hebben we vele *proefrespondenten* uitgenodigd voor deelname aan diepte-interviews en focusgroepen. Zonder het te weten hebben zij een bijdrage geleverd aan dit proefschrift. Zonder hen was het er niet gekomen.

Ook dank ik *Wim de Witte*, Hoofd van de Sector Ontwikkeling en Ondersteuning van de Divisie Sociale en Ruimtelijke Statistieken, de sector waar ik sinds de laatste reorganisatie van het CBS (september 2000) een nieuwe uitdaging heb gevonden als projectmanager van onderzoeks- en ontwikkelingsprojecten op het gebied van survey methodologie en non-respons, en *Pieter Everaers*, Directeur van de Divisie Sociale en Ruimtelijke Statistieken. Jullie maakten het mogelijk dat ik het proefschrift binnen afzienbare tijd kon afronden, met als resultaat het boek zoals dat er nu ligt.

Bovenal wil ik *Joop Hox* en *Jaak Billiet*, mijn promotoren, danken, en niet te vergeten *Edith de Leeuw*. Vanaf het allereerste begin was Jaak betrokken bij de werkzaamheden van het Vragenlab. Zo hebben we in 1993 samen een pre-test proefproject uitgevoerd over het Nederlandse en Belgische Kiezersonderzoek. Jaak, vanuit de wetenschap toonde jij altijd bijzondere interesse voor dit onderwerp. Dit heb ik altijd als een grote steun ervaren. Via Edith heb ik Joop leren kennen. Ik kende Edith al een aantal jaren. Samen hebben Edith, Joop en ik hoofdstuk 3 geschreven. Ook heeft Edith in 1996 meegewerkt aan het uittesten van de POLS-vragenlijst (Permanent Onderzoek Leefsituatie; zie hoofdstuk 8). Met plezier denk ik terug aan dat jaar, toen Edith eens per week het CBS kwam bezoeken en intrek nam in mijn kamer. En in januari 2001 werkte ze mee aan een CBS-cursus vragenlijstontwikkeling en pre-test methodes (samen met Pamela Campanelli). Haar opgewektheid en enthousiasme werkten steeds weer aanstekelijk op mij. Edith, ik dank je voor alles; ik heb altijd het gevoel gehad dat je oprecht meeleefde met de wording van mijn proefschrift. En toen jouw Joop in 1997 hoogleraar werd, heb ik hem niet lang daarna gevraagd mijn promotor te worden. Joop, jouw enthousiasme en pragmatisme hebben mij gesterkt om het proefschrift snel af te ronden. Jij hield het doel steeds strak voor ogen: zorgen dat het proefschrift klaar komt. Onze samenwerking stopt niet nu het proefschrift is afgerond. Joop heeft mij uitgenodigd om één dag per week te komen werken bij zijn vakgroep in Utrecht. Toen hij mij van zijn plannen vertelde was ik bijzonder vereerd. Met veel plezier kijk ik uit naar onze samenwerking, waarin we onze ideeën over onderzoek naar datakwaliteit verder kunnen uitwerken.

Ik dank *Harm 't Hart, Peter van der Heijden, Hanneke Houtkoop* en *Hans van der Zouwen*, die als leden van de leescommissie het manuscript hebben gelezen; met speciale dank aan Hans van der Zouwen voor het ontdoen van dit proefschrift van de laatste storende fouten.

Het werken aan dit proefschrift heeft, naast mijn volledige baan bij het CBS, veel doorzettingsvermogen gekost, waarbij de steun van de mensen om mij heen onontbeerlijk was. Het zal duidelijk zijn dat de totstandkoming van dit levenswerk (!) dan ook niet alleen mijn verdienste is.

De belangrijkste mensen in mijn leven heb ik nog niet genoemd. Begin 1996 ontmoette ik *Doortje*, toen we allebei werkten we aan een vragenlabproject over POLS (hoofdstuk 8), en in 1998 verhuisden we naar Voerendaal. Doortje, ik wil jou danken voor je geduld, dat vooral in de afrondende fase van het proefschrift op de proef werd gesteld, jouw steun, en vooral het plezier dat jij aan mijn leven geeft, steeds weer als ik jouw charmante lach zie. Jouw aanwezigheid maakt dat ik niet alleen maar met 'het werk' bezig ben. Ook *Vincent* en *Wouter*, de beide zonen van Doortje, wil ik hier noemen. Zij vroegen zich steeds weer opnieuw verwonderd af wat ik toch aan het doen was boven en of dat boek nou nog niet klaar was.

Tot slot wil ik *mijn ouders* danken voor hun onvoorwaardelijke steun en interesse in mijn leven.

Ger Snijkers
Voerendaal, oktober 2001

Summary

In the literature on questionnaire design and survey methodology, pre-testing is mentioned as a way to evaluate questionnaires (investigate whether they work as intended) and control for measurement errors (assess data quality, i.e. validity). As the American Statistical Association puts it (ASA, 1999, p. 11): “The questionnaire designer must understand the need to **pretest**, **pretest**, and then **pretest** some more.” Clark and Schober (1992, p. 29) indicate why this need to pre-test: “Surveyers cannot possibly write perfect questions, self-evident to each respondent, that never need clarification. And because they cannot, the answers will often be surprising.”

In the every-day practice of survey design, however, pre-testing and its results are not always accepted. A general feeling towards pre-testing is expressed by Converse and Presser (1986, pp. 51-52): “Pretesting a survey questionnaire is always recommended -no text in survey methods would speak against such hallowed scientific advice- but in practice it is probably often honored in the breach or the hurry. There is never the money nor, as deadlines loom, the time, to do enough of it. There is a corollary weakness that the practice is intuitive and informal. There are no general principles of good pretesting, no systematization of practice, no consensus about expectations, and we rarely leave records for each other. How a pretest was conducted, what investigators learned from it, how they redesigned their questionnaire on the basis of it – these matters are reported only sketchily in research reports, if at all. Not surprisingly, the power of pretests is sometimes exaggerated and their potentials often unrealized.”

In this Ph.D. thesis I have tried to systematically describe my experiences with pre-test research at the Questionnaire Laboratory at Statistics Netherlands, a cognitive laboratory which started its work in 1992. The purpose of the thesis is documentation of this practice. This text is not aimed at a theoretical discussion of cognitive laboratory methods, but focuses on the application of these methods: setting-up and carrying-out pre-test research, analysing the data and presentation of the results. As far as I know, in the literature on cognitive aspects of survey methodology (CASM), little is said about how to apply this kind of research in practice. This is confirmed by Willis et al. (1999, p. 137), who discuss systematic schemes for the description of how cognitive interviewing methods are practised. They conclude that “(…) no such scheme exists for use in cognitive interviewing research.” With this thesis I have tried to provide such schemes by systematically describing the state of the art at the Questionnaire Laboratory at Statistics Netherlands, in order for other researchers to continue from here. I do hope I have succeeded in this goal.

Cognitive (pre-)testing is aimed at improving the data quality, by improving the questionnaire. By means of small-scale pre-testing the questionnaire is validated, i.e. errors in the questionnaire that cause systematic errors in the question-and-answer process of the respondent in an interview setting are detected, explained and improved (in an iterative process). In this way, the

questionnaire will be adapted to the question-and-answer process and becomes easier to answer, within a shorter period of time, and will be more respondent-friendly, resulting in reduced measurement errors, i.e. increased quality (internal validity) of survey data, and reduced respondent burden. This is the CASM paradigm. The CASM paradigm defines the central hypothesis in this thesis: pre-testing in a cognitive laboratory validates questionnaires.

The question-and-answer process has been modelled by Tourangeau and Rasinski. In 1988 they presented a 4-stage model. According to this model, there are 4 stages in the question-and-answer process: comprehension of the question, information retrieval, information judgement and integration to get an answer, and reporting of that answer. The application of cognitive research to information processing in survey responding (Jobe & Mingay, 1991, p. 178), "offered a view of the survey respondent as a question-and-answer system that carried out a series of mental operations, such as comprehension of what was required in response, retrieval of relevant information from memory, and decision-making to arrive at and provide answers to the survey interviewer's inquiry." According to Jobe and Mingay (1991, p. 178), "modelling the respondent's mental operations represented a vast change over the simple stimulus-response conception of respondent behaviour, that from the beginning of modern survey-taking governed the principles employed in designing survey instruments."

The question-and-answer process is investigated in a cognitive laboratory. Pre-test research, as carried out in a cognitive laboratory, typically consists of a small number of cognitive interviews. In these interviews respondents are requested to provide information on their mental processes when answering a question in an interview. Methods to research the question-and-answer process, used at the Questionnaire Laboratory, are: expert (re-)appraisal, focus groups, in-depth interviews (including thinking aloud, follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes), and behavioural coding.

This thesis consists for a large part of papers that separately have been published before as CBS-reports, research papers, or papers presented at international conferences. These papers are based on individual research projects that have been carried out from 1992 up to 2000. For this reason, there is some overlap in the chapters and redundancy of information. In order to make this a coherent thesis, some of the papers have been slightly edited. This was done in such a way however, that every chapter can be read as a separate paper still.

The thesis starts with an introductory chapter, introducing cognitive laboratory research and the CASM paradigm. This chapter also includes the history of the CASM movement (which came about in the Unites States in the 1980's) and the history of the Questionnaire Laboratory at Statistics Netherlands, as well as an overview of the thesis. The next two chapters address aspects of computer-assisted interviewing. Since at Statistics Netherlands most questionnaires are computerised, this sets the conditions for pre-test research at the Questionnaire Laboratory. Chapter 2 discusses computer-assisted interviewing; chapter 3 addresses the effects of computer-

assisted interviewing on data quality. These three chapters are introductory to the chapters that follow, the actual core of the thesis, discussing the application of cognitive laboratory methods, including several case studies.

The methods used at the Questionnaire Laboratory at Statistics Netherlands are discussed in chapters 4 and 5. Chapter 4 presents an overview of cognitive laboratory methods. In this chapter expert appraisal, focus groups, in-depth interviewing (including follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes), and behavioural coding are discussed from a practical point of view, i.e. how they are applied in the Questionnaire Laboratory. This chapter provides an overview of current best practices. Computer-Assisted Qualitative Interviewing (CAQI) is discussed in chapter 5. The CAQI method has been developed at the Questionnaire Laboratory at Statistics Netherlands to pre-test computerised questionnaires. With CAQI a pre-test protocol is integrated in a computerised questionnaire to be tested.

In the next four chapters, case studies of cognitive research in which the methods addressed in the chapter 4 and 5 have been applied, are presented. These chapters discuss the design and the results (i.e. identified problems in the questionnaire and recommendations for improvement) of these studies. Chapters 6 and 7 are reprints of reports presented to the client, showing how the research and its results have been presented.

In the chapters 6 and 7, *laboratory* studies using CAQI in-depth interviews are described. In chapter 6 a questionnaire on income is pre-tested, using in-depth interviews with thinking aloud, follow-up probes, meaning-oriented probes and targeted test questions. Chapter 7 describes a pre-test study on questions on daily activity, pension schemes, and education and training from the European Community Household Panel. In this study in-depth interviews with thinking aloud, follow-up probes, meaning-oriented probes and vignettes have been conducted.

In chapter 8, a CAQI *field* study on the Continuous Survey on Living Conditions is described. In this qualitative operational field test respondent reaction coding and targeted test questions have been used. (In the preceding laboratory study in-depth interviews and focus groups had been conducted.)

In chapter 9 an example of a focus group study is presented. With this study a redesigned form of a business survey has been pre-tested: the Annual Establishment Production Survey. While in the other studies interviewer-administered questionnaires for persons and households were tested, here we have an example of applying cognitive methods to a *business survey* with a self-completion form. In this chapter, this study is presented within a discussion on a strategy of active respondent communication. This strategy is aimed at stimulating and motivating sample respondent to participate in surveys. The role of a cognitive laboratory in this strategy is to develop well-designed questionnaires, that look attractive and are easy to understand and to complete, thus

reducing response burden. Here, cognitive laboratory research is placed within a broader view on Total/Tailored Survey Design and Total Quality Management.

Chapter 10 concludes this thesis. This chapter provides a summary of this thesis, including a summary of methods used and results of the pre-test studies. A number of identified problems in the investigated questionnaires are: technical (jargon), vague or unclear wording, complex syntax, long question, double-barrelled question, conflict with previous question(s), question asks for specific information that is not available by heart, difficult to come to an answer because of complex calculation, overlapping or missing response items. In this chapter the identified problems are related to design errors in the questions, according to question design principles.

Although pre-testing is a way to evaluate questionnaires and control for measurement errors, in the practice of survey design, its results -including recommendations for improving the questionnaire- are not always accepted. In this last chapter, arguments against pre-test research are discussed, as well as strategies for presentation and gaining acceptance of this kind of research and its results, one of them being the application of pre-test methods according to scientific principles. To complete this thesis, the improvement of the computerisation of questionnaires is discussed by describing a limited number of tests addressing the programming process of computer-assisted survey instruments and the human-machine interaction. Finally, this thesis is concluded by discussing ideas for future research, and by showing that pre-testing in a cognitive laboratory validates questionnaires.

References

ASA, 1999, Designing a Questionnaire. ASA series: What is a Survey? (American Statistical Association, Alexandria, VA.)

Clark, H., and M. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), Questions about Questions, pp. 15-48.)

Converse, J.M., and S. Presser, 1986, Survey Questions. Handcrafting the Standardized Questionnaire. (Quantitative Applications in the Social Sciences Series, No. 63, Sage, Beverly Hills.)

Jobe, J.B., and D.J. Mingay, 1991, Cognition and Survey Measurement: History and Overview. (Applied Cognitive Psychology, Special Issue on Cognition and Survey Measurement, Vol. 5, No. 3, pp. 175-192.)

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds.), 1999, Cognition and Survey Research. (Wiley, New York.)

Tanur, J.M. (ed.), 1992, Questions about Questions: Inquiries into the Cognitive Bases of Surveys. (Russel Sage Foundation, New York.)

Tourangeau, R., and K.A. Rasinski, 1988, Cognitive Processes underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol. 103, No. 3, pp. 299-314.)

Willis, G.B., Th.J. DeMaio, and B. Harris-Kojetin, 1999, Is the Bandwagon headed to the Methodological Promised Land? Evaluating the Validity of Cognitive interviewing Techniques. (In: Sirken et al. (eds.), Cognition and Survey Research, pp. 133-153.)

Chapter 1
Introduction

The Questionnaire Laboratory at Statistics Netherlands: A Facility for Pre-testing Questionnaires

Summary: *In 1992 the Questionnaire Laboratory at Statistics Netherlands was established. This laboratory offers facilities and methods for pre-testing questionnaires, in order to reduce errors in the questionnaire. These errors are brought to light by investigating the question-asking-and-answering process with respondents, using cognitive interviewing techniques. Once these errors have been detected the questionnaire can be improved, thus getting a respondent-friendly questionnaire, reduced response burden and valid survey data. Questionnaire or cognitive laboratories originated from the movement on Cognitive Aspect of Survey Methodology (CASM), which started in the USA and Germany in the 1980's.*
This chapter is an introductory chapter to my Ph.D. thesis. The thesis is based on my work as a researcher at the Questionnaire Laboratory and is a collection of research I carried out in co-operation with colleagues; it is a documentation of my experiences. The thesis gives an overview of the state of the art of applied cognitive research at the Questionnaire Laboratory at Statistics Netherlands. This chapter discusses the CASM movement, as well as the CASM paradigm: The goals and objectives of cognitive research. Also the history of the Questionnaire Laboratory and an overview of the thesis will be presented.

Keywords: *Questionnaire or Cognitive Laboratory, Cognitive Aspects of Survey Methodology, Question-and-Answer Process, Data Quality, Measurement Errors, Validity, Improving Questionnaires, Response Burden, Overview of the Thesis.*

1.1. Introduction

The Questionnaire Laboratory at Statistics Netherlands started its work in 1992. This laboratory offers facilities and methods for (pre-)testing and developing questionnaires. A (pre-)test program typically consists of a small number of cognitive interviews. In these interviews respondents are requested to provide information on their mental processes when answering a question in an interview by using cognitive interviewing techniques like thinking-aloud and probing.

Questionnaire or cognitive laboratories originated from the movement on Cognitive Aspect of Survey Methodology (CASM), which started in the USA and in Germany in the early 1980's. This movement studies cognitive aspects related to the questionnaire by investigating the question-asking-and-answering or question-and-answer process of respondents. These aspects are investigated within a cognitive laboratory. The objectives of the cognitive laboratory are to improve the quality of survey data and reduce response burden, by validating the questionnaire.

This chapter is an introductory chapter to my Ph.D. thesis. The thesis is based on my work as a researcher at the Questionnaire Laboratory and is a collection of research I carried out in co-operation with colleagues; it is a documentation of my experiences. I started with the Questionnaire Laboratory in early 1992, as a senior researcher. From 1995 up till 2000 I was a senior researcher and project leader of advisory and research projects. In 2000, I left the Questionnaire Laboratory to become a manager of research and development projects in social statistics. The thesis gives an overview of the state of the art of applied cognitive research at the Questionnaire Laboratory at Statistics Netherlands.

A brief history of the CASM movement is discussed in the next section of this introductory chapter. Section 1.3 discusses the CASM paradigm: the goals and objectives of cognitive research. In the last section an overview of the thesis will be presented from an historical perspective (the history of the Questionnaire Laboratory at Statistics Netherlands), as well as a systematic overview of the topics to be discussed in the following chapters.

1.2. Cognitive Aspects of Survey Methodology

Questionnaire or cognitive laboratories fit within a movement in survey methodology which came about in the United States in the 1980's to study Cognitive Aspects of Survey Methodology (CASM). The history of CASM is described in detail by Jobe and Mingay (1991), Tanur and Fienberg (1992), Sudman, Bradburn and Schwarz (1996), and Tanur (1999).

In the United States during the 1970's, non-response rates to surveys rose. Because of this, there was an increasing concern during those years about the validity of the survey data on which government policy and academic research was being based. In the U.S., as well as in the U.K., a

series of conferences were held to suggest ways to increase the validity of survey data, and thus the validity of derived conclusions. These conferences were the roots of the CASM movement.

As Tanur and Fienberg (1992), and Tanur (1999) report, the 'official' beginning of the CASM movement is usually dated as June 1983 in the United States and July 1984 in Germany. Under the auspices of the Committee on National Statistics of the U.S. National Academy of Sciences, an Advanced Research Seminar on Cognitive Aspects of Survey Methodology was organised. The name of the seminar was chosen deliberately so that the resulting acronym, CASM, would signify the deep interdisciplinary 'chasms' that would have to be bridged if the proposed collaborative efforts were to be successful. In this seminar, statisticians, survey researchers, cognitive psychologists, anthropologists, and U.S. government agency staff participated. Apart from addressing the problems arising in surveys and how cognitive theories and methods might be applied toward their solutions, they also addresses how surveys might be used by those in the cognitive sciences to expand beyond the walls of their laboratories. Jabine, Straf, Tanur and Tourangeau (1984) presented a report of the seminar. A parallel development, a conference on Social Information Processing and Survey Methodology at ZUMA (Zentrum für Umfragen, Methoden und Analysen) in Mannheim, Germany, addressed many of the same issues (Hippler, Schwarz & Sudman, 1987). At the same time in the U.K., Belson carried out some interesting research to investigate the understanding of survey questions and the validity in survey research, using techniques like rephrasing questions and 'thinking aloud' in retrospective "intensive individual interviews" (Belson, 1981, 1986).

According to Jobe and Mingay (1991), the CASM seminar incorporated two features that set the stage for future developments in highly specific ways. One was the videotaping and subsequent critical analysis of an actual interview performed in the routine course of the U.S. National Health Interview Survey. The videotaped interview proved to be a provocative tool in the generation of ideas for future research. The other was the detailing of several specific research undertakings to be put into practice. These undertakings were aimed at producing examples of what achievements might be expected from a large-scale, interdisciplinary research programme on the cognitive aspects of survey methodology. Included in those ideas were, as Jobe and Mingay (1991, p. 177) report, "a series of laboratory studies to explore cognitive processes in survey responding, to investigate whether protocol analysis (cf. Ericsson & Simon, 1984) would be an effective method of studying how respondents retrieve information from memory and answer survey questions, to investigate how the feelings associated with the response to one attitude question may influence responses to later questions, and to investigate cognitive processes and knowledge representations that are implicit in responses to many survey questions (see Jabine et al., 1984). All but one of these specific research projects were subsequently supported, mostly by the National Science Foundation."

The application of cognitive research to information processing in survey responding (Jobe & Mingay, 1991, p. 178), "offered a view of the survey respondent as a question-and-answer system

that carried out a series of mental operations, such as comprehension of what was required in response, retrieval of relevant information from memory, and decision-making to arrive at and provide answers to the survey interviewer's inquiry." According to Jobe and Mingay (1991, p. 178), "modelling the respondent's mental operations represented a vast change over the simple stimulus-response conception of respondent behaviour, that from the beginning of modern survey-taking governed the principles employed in designing survey instruments." (See figure 1.1 in section 1.3; see also Snijkers & Luppes, 2000, chapter 9).

One major benefit to survey research of the collaboration with cognitive sciences is, according to Jobe and Mingay (1991, p. 184), "the opportunity it provides to test and evaluate questionnaires using cognitive interviewing methods." Since the CASM seminar, a series of government laboratories were set up to explore cognitive aspects of surveys. Tanur (1999) provides a list of these laboratories. The first laboratory was established by Monroe Sirken at the U.S. National Center for Health Statistics (NCHS): the Questionnaire Design Research Laboratory, which conducts applied cognitive research in developing, designing, and pre-testing survey questionnaires. In 1985 the QDRC at Statistics Canada was established. In 1988 the U.S. Bureau of Labor Statistics established a cognitive laboratory (Dippo & Norwood, 1992), shortly after followed by the U.S. Census Bureau. In Europe, cognitive work does not enjoy such formal status, but is carried out, among other places, by researchers at the national statistical agencies of The Netherlands, Sweden and the UK, and ZUMA. Other major benefits for survey research are, according to Jobe and Mingay (1991), the use of cognitive methodology to conduct experimental laboratory research (in stead of or prior to field experiments) on issues of interest to survey researchers, and the application of cognitive methods and theory to survey experiments conducted in the field.

In the 1980's a lot of research has been done in the field of cognition and survey research. Jobe and Mingay (1991, pp. 180-181) present a long list of papers clustered by topic: Autobiographical frequency and magnitude estimation; Cognitive interviewing techniques; Experimental memory; Influence of response alternatives on responses; Memory for dates and events; Question comprehension; Question order/question wording; and Response order.

During the 1990's "the CASM movement grew and prospered" (Tanur, 1999, p. 16). Survey research increasingly applied cognitive methods and took for granted that cognitive pre-testing was essential for any serious survey enterprise. At that time several conferences on the movement had been organised, resulting in several publications, see e.g. Tanur (1992), Schwarz & Sudman (1992, 1994, 1996), and Sudman, Bradburn & Schwarz (1996). Also in conferences not especially focussed on the movement, the application of cognitive pre-testing of questionnaires was discussed (see e.g. Biemer et al., 1991; Cox et al., 1995; Lyberg et al., 1997).

However, after the 10th anniversary of CASM in 1993, staff at the U.S. National Center for Health Statistics realised that there was no obvious driving force for the movement anymore. "Nor

had there been any recent evaluation of the movement's progress or prospects. Nor, was there any strong impetus to incorporate other disciplines into the CASM dialogue or to examine cognitively the other phases of the survey process" (Tanur, 1999, p. 16). Hence, a Second Advanced Research Seminar in the Cognitive Aspects of Survey Methodology (CASM II) was organised in 1997, resulting in two publications. The proceedings (Sirken, Jabine et al., 1999) summarised the history of the CASM movement, reviewed current needs and proposed future directions for interdisciplinary survey methods research. The second publication is 'Cognition and Survey Research' (edited by Sirken, Herrmann et al., 1999). In the CASM II seminar, leading survey researchers, cognitive psychologists, and other scientist critically reviewed the impact of CASM research since 1984 and discussed the important roles of computer sciences, statistics, and other scientific disciplines in a rapidly evolving field of interdisciplinary survey methods research.

1.3. The CASM paradigm

Cognitive research applied to survey methodology originated from the need to improve the quality of survey data, and in particular the validity of data. One way to do that, is, as we have seen, pre-testing the questionnaire (before conveying the survey in the field) with respect to information processing in survey responding. This is what cognitive laboratories are focussed at.

The quality of survey data is determined by a number of aspects of the survey design, as sources of observational errors[1] in survey data collection (Groves, 1989; Snijkers, 1992, see chapter 2, figure 2.1; Akkerboom, 1992). These are:

- the questionnaire (including, respondent instructions and task difficulty; question wording; question structure like length of question, open or closed, number of response options, 'don't know' option, sensitive topic, recalling past events; order of the questions and position of the question in the questionnaire);
- the interviewer (training, experience, 'interviewing style' and other personal characteristics);
- the respondent (willingness and motivation to participate, interest in subject of the survey, education level or cognitive skills and other personal characteristics);
- the interviewing mode (interviewer-administered personal interview, telephone, or self-administered)
- the interviewing technique (paper-and-pencil or computer-assisted);
- the interview situation (the interaction of the five aspects mentioned above, e.g. the affection between interviewer and respondent; disturbances during the interview).

[1] Apart from observational errors ("deviations of the answer of respondents from their true values of measurement"), Groves (1989, p. 11) defines non-observational errors: "errors arising because measurements were not taken on part of the whole population." Non-observational errors (coverage, non-response and sampling errors) are not discussed here.

Cognitive laboratory research is focussed at reducing observational or measurement errors associated with these aspects of survey data collection. However, the focus is especially on the questionnaire, the respondent and their interaction during the course of an interview.

These measurement errors may be systematic or random, causing survey data to be not valid as well as not reliable. This results in biased estimates of population estimates or increased variance of those estimates. Testing questionnaires in a cognitive laboratory is aimed at detecting and reducing systematic errors in the question-asking-and-answering process (i.e. validating the questionnaire) caused by errors in the questionnaire. By reducing these errors, the validity – more in particular the internal validity[2] – of survey data is increased.

The question-and-answer process has been described by Tourangeau and Rasinski (1988). They distinguish four phases: (1) interpretation and comprehension, (2) information retrieval, (3) judgement, and (4) reporting (see figure 1.1). In the first step (1), a respondent has to interpret the question, i.e. interpretation and comprehension of the question and the response task. As for the question, comprehension is dependent on wording[3], syntax, other aspects of the question (like lengths of the question, or double-barrelled question), and reference frame (set by e.g. previous questions[4], loaded terms[5], or response options[6]). Comprehension of the task has to do with understanding what the questions asks the respondents to do, in order to provide an adequate answer, i.e. a preference ordering of items. This step may results in several conclusions: the question is perfectly clear to the respondent, the question is not clear (because of question wording, task difficulty) or the respondent thinks the question is clear to him but in fact it is not[7].

In the second step (2) the respondent has to retrieve information from his memory (information retrieval). In case the information cannot be retrieved from memory other sources may have to be consulted. Information retrieval of information is dependent on the information that is asked for

---

[2] 'Internal' refers to "the set of respondents examined in the data collection, not to others outside the set of subjects, nor to situations outside the data collection itself. Threats to internal validity (…) include problems of operationalization." (Groves, 1989, p. 27; see also footnote 1.)

[3] Clark and Schober (1992) discuss general principles regarding understanding of questions and responding, based on language use in conversations. As for wording they state (pp. 27-28): "Respondents answer vaguely worded questions in idiosyncratic ways."

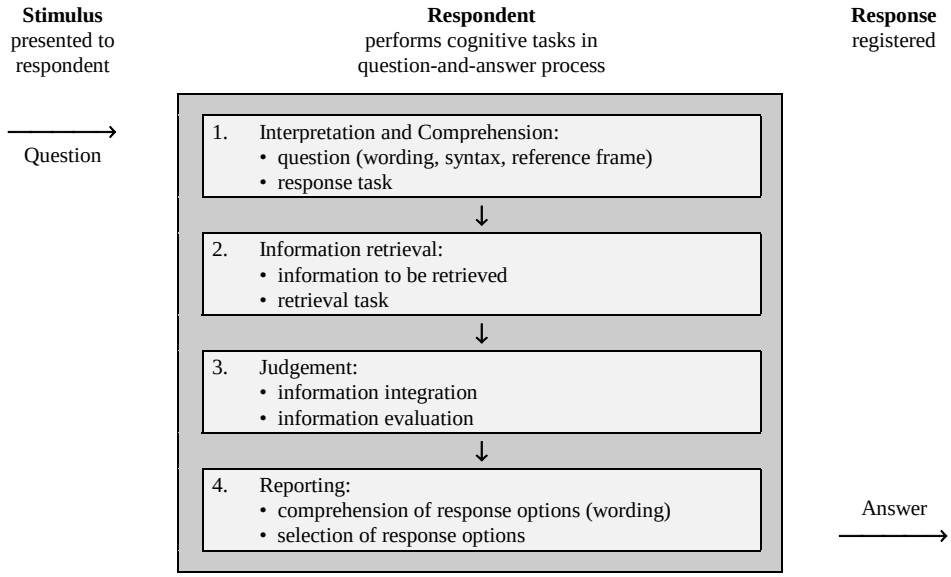
[4] As with regard to question sequence, Clark and Schober (1992, see footnote 3) state:
- "People interpret successive questions as related in topic – unless they are told otherwise." (p. 40)
- "When a general question follows a specific question on the same topic, it may get an exclusive or an inclusive interpretation, depending on the circumstances." (pp. 40-42)

[5] "Loaded terms help set the perspective from which questions are to be answered." (Clark & Schober, 1992, see footnote 3, pp. 29-30.)

[6] As for response options Clark and Schober (1992, see footnote 3, pp. 30-34) state:
- "The response alternatives to a questions help determine the domain of inquiry in which is to be answered."
- "Questions with and without response alternatives imply different perspectives."
- "Response alternatives are often construed as what is typical or normal in the population."

[7] "Respondents fail to see when the surveyer is using words differently from the way they use them." (Clark & Schober, 1992, see footnote 3, pp. 28.)

Figure 1.1. The question-and-answer process within the stimulus-response model of survey responding



and the retrieval task. According to Tourangeau and Rasinski (p. 300) it is unlikely that respondents retrieve all relevant information about an issue. Therefore, this phase can be seen as "a kind of sampling process that over-represents the most accessible beliefs or situation cues". For example, the question may be asking for specific[8], past, or proxy information. The retrieval task is the task the respondent has to carry out, in order to get the information needed to come to an answer. This task may be difficult because other persons or records may have to be consulted.

After this, the respondent has to come to an answer in the third step (3): the judgement or evaluation phase. In this phase all retrieved information has to be combined into an overall judgement or 'internal' answer. Also, in this phase, the 'internal' answer may be checked for consistency with prior answers[9] or for social desirability. Sometimes this step is trivial. In most cases, however as Tourangeau and Rasinski point out (p. 300), "the question will not map directly onto an existing belief or a summary evaluation, and a complicated process will be needed to generate a judgement answer from the set of retrieved beliefs." For example, the answer has to be guessed, or a complex calculation has to be made[10]. Step 2 and 3 may be taken together as the information processing phase.

[8] "Respondents seem to have stable opinions on issues they know nothing about." (Clark & Schober, 1992, see footnote 3, pp. 28-29.)
[9] "People use their earlier answers as evidence for their later judgments." (Clark & Schober, 1992, see footnote 3, pp. 37-39.)
[10] "Respondents estimate factual answers that would take too long to figure out." (Clark & Schober, 1992, see footnote 3, pp. 35-36.)

In the final step (4), respondents must report their answer. In this reporting stage respondents must map their 'internal' answer onto one of the response options, in case of closed questions, or the response format in open questions, i.e. the unity asked for in the questions (e.g. meters or inches). According to Tourangeau and Rasinski, the final response given may be a compromise between the respondent's judgement and the predefined response options. The judgement may be changed when no option seems appropriate, i.e. when options are missing or overlapping. Reporting is also dependent on the comprehension of the response options (wording).

This model is the theoretical basis for cognitive laboratory research, as we have seen in section 1.2. Methods to be used for testing questionnaires and analysing cognitive interviews are based on this model. In in-depth interviews e.g., respondents are requested to think aloud while answering a question. Also probes may be used to focus the attention to one specific aspect of the question, like the interpretation of a specific term in the question. Chapter 4 provides an overview of cognitive laboratory methods. By applying these methods, problems in the question-and-answer process are revealed.

As we have seen above (figure 1.1), problems related to comprehension have to do with the question wording, the syntax of the question, the question itself (e.g. long or double-barrelled question), the reference set, and the response task. Problems with retrieval of information have to do with difficulties in recalling or recognising information, and the retrieval task. Judgement problems have to do with difficulties in integration and evaluation of information. And problems in the reporting phase have to do with comprehension of the response options (wording) and selection of the adequate option(s). In the chapters 6, 7, 8 and 9 of the thesis, examples of cognitive research and their findings, i.e. identified problems, will be discussed.

Apart from problems in the question-and-answer process, also other problems may be revealed with this kind of research. These problems may again be related to the questionnaire, as well as to the other aspects of data collection that determine the quality of survey data (see above). For example, incorrect routings in the (computerised) questionnaire, flaws in the lay-out of a (self-completion) questionnaire (e.g. a paper form), ineffective or inefficient interviewer procedures, or inappropriate mode of the data collection.

Once these problems have been detected, they have to be prevented from occurring. To do that, it is necessary to explain why they occur, i.e. to identify what aspects (or errors) in the questionnaire cause these problems. The findings of cognitive interviews not only consist of identified problems in the question-and-answer process but also of indications as to why these problems emerge. Like the identified problems, these indications are also based on respondent reactions in cognitive interviews.

The next step, then, is repairing the questionnaire (c.q. the interviewer procedures or interviewing mode) by revising it (e.g. changing question wording, question order, routing, lay-out) according to the identified errors in the questionnaire. Thus, the questionnaire is adapted to the question-and-answer process. As a result the questionnaire will be easier for respondents to understand and complete, and thus becomes more respondent-friendly. This will reduce the time to complete an interview and it will be more pleasant for both the respondent and the interviewer. By improving the questionnaire in this way, systematic errors in the question-and-answer process are reduced or prevented from occurring, thus reducing or preventing related measurement errors to occur. The result of this process is a valid, respondent-friendly questionnaire, which, in its turn, results in increased internal validity of survey data and reduced response burden.

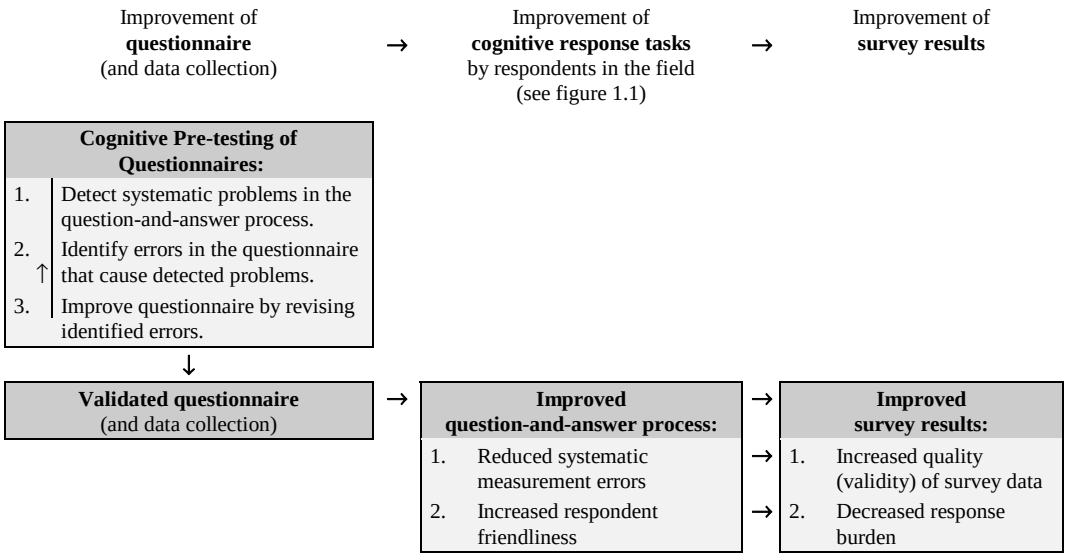
Preferably, cognitive methods are applied to investigate the questionnaire prior to collecting the survey data, i.e. *pre*-testing the questionnaire. This process of pre-testing and improving the questionnaire is an essential step within the process of developing a questionnaire. This process may be seen as an iterative process in which prototypes of a questionnaire are pre-tested, improved and pre-tested again.

However, cognitive testing of questionnaires may also be done afterwards, e.g. in the case of an ongoing survey. Then, the goal of testing is mainly to identify measurement errors. This may be helpful in explaining unexpected effects within survey data. Here, we have to keep in mind that cognitive testing of questionnaires is a small-scale, qualitative type of research, and results cannot be generalised straight away. However, on the basis of cognitive research, field tests and embedded experiments may be carried out, in order to quantify the measurement error effects (see chapter 9). Also re-interviews may be conducted, as well as debriefing interviewers.

Field tests and embedded experiments may also play a role in improving the survey design. Tanur and Fienberg (1992) state that within the iterative process of developing questionnaires "research findings about survey content and measurement techniques require careful implementation and evaluation as they are shifted from the laboratory to the field." (…) "The development of mechanisms for carrying out embedded experiments involving proposed survey improvements on a regular basis would do much to move us in the right direction. The cognitive laboratory provides a key link in this improvement process by offering a setting and access to theories for identifying alternatives for the field." In chapter 4 a model for survey design development and improvement is discussed: the 5-step (pre-test) model for data collection development (see also subsection 1.4.1). This model offers a methodology for testing all aspect of data collection carefully, shifting research findings from the laboratory to the field.

To summarise, the goals of pre-testing questionnaires in a cognitive laboratory are: (1) detection of systematic problems in the question-and-answer process, (2) identification of errors in the questionnaire that cause those problems in the question-and-answer process, and (3) improvement of the questionnaire by revising those errors. As a result of this (iterative) process,

Figure 1.2. The CASM paradigm:
Validating questionnaires and improving survey results by cognitive pre-testing



the questionnaire is adapted to the question-and-answer process: the questionnaire is validated. Thus, measurement errors are reduced or prevented to occur, and the quality of derived survey data is improved. This is the main objective of this kind of research. Also, the questionnaire will be easier for respondents to answer, and taking less time to complete. Thus, and this is the second objective, the response burden is decreased. This theory on validating questionnaires and improvement of survey results by pre-testing is the CASM paradigm (see figure 1.2).

The CASM paradigm defines the central hypothesis in this thesis: pre-testing in a cognitive laboratory validates questionnaires. The methods to be applied in pre-testing questionnaires will be described in the chapters to come, as well as how they should be applied (according to scientific standards). Also four examples of this kind of research (case studies) will be described, including their findings (i.e. identified problems in the question-and-answer process with questions), and suggestions for improvements to the questionnaires. The results of these studies support the hypothesis, since the identified problems are related to *design* errors in the questions, according to scientific standards on question design. The next section provides an overview of the issues that will be addressed in this thesis.

1.4. Overview of the thesis

This thesis is based on my work as a researcher at the Questionnaire Laboratory at Statistics Netherlands. This text is a collection of research I carried out in co-operation with colleagues; it is a documentation of my experiences. In this section an overview of the thesis will be presented. In the first subsection, the history of the Questionnaire Laboratory will be described, showing the development of those experiences, and placing the topics to be discussed in this thesis in historical context. Subsection 1.4.2 presents an overview of these topics in order of the chapters to come.

1.4.1. The Questionnaire Laboratory at Statistics Netherlands

The Questionnaire Laboratory at Statistics Netherlands was established in 1992. The immediate cause for this decision was the visit by Keller (at that time director at Statistics Netherlands) and Kooiman (at that time head of the Department of Statistical Methods) to the ISI[11] Conference in Caïro, September 1991. Here, they attended a presentation by Allen Gower (Head of the Questionnaire Design Recource Centre, QDRC, at Statistics Canada) on the cognitive laboratory of Statistics Canada (Gower, 1991, pp. 1-2):

"In Canada the QDRC was established in 1985 as a focal point for questionnaire design." (…) "The original goals of the QDRC were to become a centre of expertise in questionnaire design, to undertake research, and to provide a forum for the collection, development, and dissemination of knowledge in the field of questionnaire design." (…) "The QDRC initiated the concept of designing 'respondent-friendly' questionnaires that are easy for respondents to understand and complete. Respondent-friendly questionnaires can have beneficial impacts on the data collection process in terms of improved respondent relations, co-operation, and data quality (higher response rates and more accurately completed questionnaires) as well as reduced response burden and costs (reduced time spent on follow-ups and dealing with errors or missing information)." (…) "In providing questionnaire design support, the major thrust of the QDRC's work is research related to the development, testing, and evaluation of survey questionnaires using cognitive and qualitative methods." (…) "Focus groups, in-depth and think-aloud interviews, paraphrasing, and the observation of respondent behaviour are the most frequently used techniques."

Keller and Kooiman decided that at Statistics Netherlands also a facility for pre-testing of questionnaires had to be established. The Questionnaire Laboratory was placed within the Department of Statistical Methods in Heerlen[12], where the Department for Data Collection for Social Surveys (personal and household surveys) is also located. At the end of 1991 Hans

[11] International Statistical Institute.
[12] Statistics Netherlands has two locations: Heerlen (near Maastricht) and Voorburg (near The Hague), about 250 km (about 150 miles) apart.

Akkerboom and myself received the assignment to set up a Questionnaire Laboratory at Statistics Netherlands (Akkerboom, 1992). We started off by investigating the literature on the subject and by visiting conferences. The first conference we visited was the international workshop 'Cognition and Survey Methodology' ('t Hart & Kox, 1991) in Utrecht, November 1991[13]. This workshop came for us at the right moment, and at the beginning of 1992 the Questionnaire Laboratory was established.

At the beginning of 1993 our group was extended with two additional colleagues: Martin Luppes and Josée Nollen-Dijcks. During the building phase of the Questionnaire Laboratory (1992-1993), several tasks had to be completed before the Laboratory was in action (Akkerboom, 1996):

- training in cognitive interviewing methods[14] and developing standard techniques for quickly and efficiently carrying out cognitive interviews within Blaise,
- training in analysing cognitive interviews and developing ways to do so quickly and efficiently (Snijkers, Akkerboom & Nollen-Dijcks, 1994), and
- recruiting and registering voluntary respondents to be invited for cognitive interviews at Statistics Netherlands (Nollen-Dijcks, 1993).

In 1993 we were ready to start our first projects in questionnaire pre-testing. One of these projects was in collaboration with Billiet, whom we had met at the 'Cognition and Survey Methodology' workshop in Utrecht[15]. This project was on testing questions from the Dutch and Belgian Election Survey (Snijkers, Akkerboom & Nollen-Dijcks, 1994; Snijkers, 1994a). Other projects were on the Survey on Safety (Nollen-Dijcks & Snijkers, 1994; Nollen-Dijcks, 1994b) and advance letters (Luppes, 1994). The goals of these projects were to get experience with questionnaire laboratory research: (1) selecting survey questions for cognitive interviewing, (2) selecting the most appropriate cognitive interviewing techniques, (3) combining the two into one protocol for conducting cognitive interviews, (4) conducting the cognitive interviews, and

[13] In Utrecht we met Judith Tanur, who was about to finish a book a the subject (Tanur, 1992). We also talked to Judith Lessler (Research Triangle Institute, USA), who presented a paper on 'Cognitive Laboratory Methods' (Forsyth & Lessler, 1991a), and Jaak Billiet (Catholic University of Leuven) (Billiet, 1991).

[14] In September 1992, I visited the First International Conference Organised by the Study Group on Computers in Survey Analysis 'Survey and Statistical Computing' in Bristol, U.K. (Snijkers, 1992). Here, I met Pamela Campanelli, who was an expert in cognitive testing of questionnaires at the US Census Bureau and at that time working for SCPR (Social & Community Planning Research) in London. At the beginning of 1993 we invited her to visit Statistics Netherlands to present a short course in cognitive interviewing. In September 1993 Nollen-Dijcks visited several questionnaire laboratories in the United States: Bureau of Labor Statistics, National Center for Health Statistics, Census Bureau and Research Triangle Institute (Nollen-Dijcks, 1994a). In October 1993 Luppes visited the *Statistisches Bundesamt* in Wiesbaden, and the *Zentrum für Umfragen, Methoden und Analysen* (ZUMA) in Mannheim, Germany (Luppes, 1993). At the beginning of 1994 we also visited the Catholic University of Nijmegen for a one-day course in qualitative interviewing (Luppes, Nollen-Dijcks, Akkerboom & Snijkers, 1994). Apart from visiting other institutions and following training sessions, we also studied the literature on cognitive interviewing techniques (e.g. Converse & Presser, 1986; Fowler & Mangione, 1990; Forsyth & Lessler, 1991b; Tanur, 1992).

[15] See footnote 13.

(5) analysing the interviews. The results of these projects, including our experiences, were presented in reports.

As for the cognitive interviewing techniques, we started with thinking-aloud, follow-up and meaning-oriented probing, and paraphrasing. (An overview of methods is presented in chapter 4.) Since at Statistics Netherlands almost all personal and household surveys are carried out by means of computer-assisted personal or telephone interviewing (CAPI or CATI) using the interviewing program 'Blaise', we faced the problem to apply these methods to pre-testing computerised questionnaires: the pre-testing methods had to be adapted to computer-assisted interviewing (CAI). (Issues of CAI are discussed in chapter 2 (Snijkers, 1992) and chapter 3 (De Leeuw, Hox & Snijkers, 1995, 1998)). Within this setting, we designed a computer-assisted method for pre-testing computerised questionnaires by adding cognitive interviewing methods to the survey questionnaire. We called this method 'Computer-Assisted Qualitative Interviewing' (CAQI, which will be discussed in chapter 5; Snijkers, 1997a).

In 1994 we also started doing focus groups[16] (Nollen-Dijcks, 1995; see chapter 4). With focus groups, forms for business surveys were tested (Huigen, Nollen-Dijcks & Akkerboom, 1994). So, now we were not only testing social survey questionnaires but also questionnaires for business surveys.

At the end of 1994, Statistics Netherlands had gone through a major reorganisation, which meant that the Questionnaire Laboratory was moved to the newly formed Department of Data Collection Methodology within the new Division of Data Collection. The reorganisation caused some changes within the group. In 1995, Questionnaire Laboratory staff consisted of 9 persons: 1 manager, responsible for acquiring new projects and co-ordination of projects; 3 senior researchers (with a background in survey methodology) and 3 junior researchers (with a background in survey methodology and psychology), who carried out the testing project. Two assistants for recruiting and selecting voluntary respondents, preparing the CAQI questionnaire in Blaise, and making other arrangements, assisted them.

The years to come (up to 1998) were the most productive ones[17]. Every year, five to ten questionnaires (both for personal and household surveys and business surveys) were pre-tested by means of cognitive interviewing methods. The testing projects carried out were mostly commissioned by internal clients. It took 1 to 3 months to finish those projects, depending on the wishes of the client: when the results should be ready, the number of questions to be tested, the

[16] Early 1995, we invited Helen Finch (SCPR, London) for a short course in focus group techniques.
[17] During those years a lot of conferences were visited to present papers on pre-testing projects (Snijkers, 1995a, 1996a, 1996b, 1998; Snijkers, Akkerboom, Kuipers & De Leeuw, 1996; Snijkers, De Leeuw, Hoezen & Kuipers, 1999a). In March 1997, I was invited by the US Census Bureau (Jennifer Rothgeb) and Statistics Canada (Allen Gower) to discuss our experiences with CAQI (Snijkers, 1997b).

number of interviews, the detail of the report (briefly or with detailed verbal transcripts). The number of in-depth interviews, with voluntary respondents, varied between 5 (for a quick scan) to 50 (with various groups of respondents). The number of focus groups varied between 1 to 5, each group consisted of 6 to 12 participants. The participants could be respondents, field interviewers or researchers. Those interviews and focus groups were carried out in a room with a one-way mirror, behind which an observer watched the interviews and made notes. Usually, the interviews were also audio taped, and sometimes even recorded on video. Examples of this kind of research are presented in chapters 6 (on income: Snijkers, 1994b) and 7 (questions from the European Community Household Panel: Snijkers, 1995b, 1995c; Van de Donk & Snijkers, 1995).

In addition to this kind of elaborate research, also expert reviews were done, without conducting cognitive interviews. Often the Questionnaire Laboratory was asked to review an advance letter or a questionnaire at the last minute of the preparation stage of a survey. Then, 1 to 3 researchers commented upon the letter or the questionnaire within a few days. The researchers would discuss their comments and come to one conclusion (see chapter 4).

During those years, the methods used were extended with other methods, like qualitative field tests in which cognitive interviewing methods like probing and reaction coding were being applied. These methods were not only used to test the questionnaire in the field, but also to test the data collection method as a whole (including interviewer procedures and interviewing mode). In 1995 a major redesign program was set up to combine several surveys on living conditions into one Continuous Survey on Living Conditions (POLS: Bakker & Winkels, 1998). The newly designed questionnaire was first tested in the laboratory and then in the field (De Heer, 1996; Dehue, 1996; Snijkers, 1996c; Akkerboom, Dehue & Snijkers, 1998). In chapter 8 (Snijkers, De Leeuw, Hoezen & Kuipers, 1999b) this field test will be described.

After several years of doing research, gradually, a systematic way of testing questionnaires and data collection procedures had been developed. These methods and techniques were put together into a model for developing and testing questionnaires and survey data collection procedures: The 5-step (pre-)test model for data collection development (Akkerboom & Luiten, 1996; Snijkers, 1996b; Akkerboom & Dehue, 1997; Dehue, 1997; Akkerboom, Dehue & Snijkers, 1998; Snijkers, 1998). In 5 steps the data collection process is developed and pre-tested, gradually shifting from testing a first draft of the questionnaire to testing the data collection procedure in the field, and implementation of the survey. This model offers a methodology to shift research findings “from the laboratory to the field” (Tanur & Fienberg, 1992, see section 1.3).

The first step in the model is a definition study, in which the questionnaire and the data collection procedure are developed and reviewed. In the second step cognitive interviews are conducted in order to pre-test the questionnaire. Ideally, in this phase the questionnaire is improved in an iterative process of testing and improving. The third step is a qualitative field test in which the improved questionnaire and data collection procedures are tested in the field. In the fourth step,

a large-scale quantitative pilot study is carried out, resulting in the final questionnaire and data collection procedures. The fifth step is the implementation of the survey. This model will be discussed in chapter 4, including the cognitive research methods that can be used in every step.

Since 1998, in addition to pre-testing questionnaires, attention of the Questionnaire Laboratory was also given to developing new data collection methods, including computer-assisted web interviewing (see chapter 3), and procedures to reduce non-response. At Statistics Netherlands non-response was considered a major problem, both for social and business surveys. In 1999 a program was started to generate ideas on how to increase response rates for business surveys by improving the communication with the respondent. The traditional approaches based on simple communication strategies (one stimulus for all units at the same moment, traditional reminder approaches using authority principles) give serious problems with response time, net response, and response quality. This program was aimed at developing measures that change traditional, formal and passive contact strategies into active, respondent driven and motivational approaches. Within this approach, well-designed, and tested respondent-friendly questionnaires are essential, in order to reduce response burden. Thus, cognitive laboratory research is related to Total/Tailored Survey Design and Total Quality Management. This approach will be discussed in chapter 9 (Snijkers & Luppes, 2000).

In 2000 Statistics Netherlands was reorganised again. The Questionnaire Laboratory was moved back to the Department of Statistical Methods, where it continues to do its work as a facility for pre-testing questionnaires. In order to improve the quality of statistics, pre-testing of questionnaires was considered important. However, since the beginning of the Questionnaire Laboratory, this methodology has also been criticised. This criticism, as well as strategies for presentation and gaining acceptance of pre-test research and its results will be discussed in chapter 10 (section 10.4).

1.4.2. Overview of the thesis: Chapter by chapter

As mentioned above, this thesis is a documentation of my experiences (including those of colleagues) at the Questionnaire Laboratory at Statistics Netherlands. It is not aimed at discussing cognitive research at a theoretical level, but from a practical point of view. This text discusses the application of cognitive laboratory methods: setting-up and carrying-out pre-test research, analysing the data and presentation of results. To my knowledge, little is said about how to apply this kind of research in practice in the literature on cognitive aspects of survey methodology (cf. Willis et al., 1999). When we started with the Questionnaire Laboratory at Statistics Netherlands we had to find this out for ourselves. With this text I will give a systematic overview of the state of the art of applied cognitive research at the Questionnaire Laboratory at Statistics Netherlands, in order for other researchers not to start from scratch again.

This text consists for a large part of papers that separately have been published before as CBS-reports, research papers or papers presented at international conferences. For this reason, there is some overlap in the chapters and redundancy of information. In order to make this a coherent thesis, some of the papers have been slightly edited. This was done in such a way however, that every chapter can be read as a separate paper still. The thesis started with this introductory chapter.

1. In this chapter, I introduced cognitive laboratory research. I discussed the CASM movement, from which cognitive laboratories originated, including the history of the Questionnaire Laboratory at Statistics Netherlands. Here, I also discussed the goals and objectives of cognitive research (the CASM paradigm): Validating questionnaires and improving survey data by cognitive pre-testing.

The next two chapters address aspects of computer-assisted interviewing. Since at Statistics Netherlands most questionnaires are computerised, this sets the conditions for pre-test research at the Questionnaire Laboratory. Within this setting, we developed a method to pre-test computerised questionnaires: Computer-Assisted Qualitative Interviewing (CAQI, chapter 5).

2. Chapter 2 discusses computer-assisted interviewing (CAI). The focus in this chapter is on comparing computer-assisted telephone to personal interviewing (CATI and CAPI), with regard to characteristics of these modes, costs and data quality. The discussion also includes general characteristics of CAI (like automated routing, tailoring of question wording, range and consistency checks). These characteristics also hold for CAQI. (This chapter is a reprint of Snijkers, 1992.)
3. Chapter 3 addresses the effects of computer-assisted interviewing on data quality. Since we also use the computer to conduct cognitive interviews (in CAQI), aspects of data quality related to the use of the computer will be discussed here, including respondent and interviewer attitudes towards the computer. The main conclusions are that both respondents and interviewers accept computer-assisted data collection methods, and that survey data quality improves, especially when complex questionnaires are used. In general, respondents are positive about the use of the computer during an interview: they attribute a greater degree of professionalism to the interview. The social interaction with the interviewer is described as comfortable and relaxed. As for the interviewers, the computer makes additional interviewer training in computer usage and computer-assisted interviewing necessary. (This chapter is a reprint of De Leeuw, Hox & Snijkers, 1995, 1998.)

Chapters 4 and 5 discuss the methods that are applied in the case studies that follow (chapters 6, 7, 8 and 9).

4. Chapter 4 presents an overview of methods to be used in cognitive laboratory research. In this chapter expert appraisal, focus groups, in-depth interviewing (including follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes), and behavioural coding will be discussed from a practical point of view, i.e. how they are applied in the Questionnaire Laboratory. This chapter provides an overview of current best practices. These methods are presented within the context of the 5-step (pre-)test model for survey data

collection development. For each step in the process of developing a data collection procedure, the model indicates what methods can be used. In a full pre-test program, following this model and the guidelines (including scientific standards) for application discussed in this chapter, all aspects of the survey will be carefully tested in advance. This model offers a methodology to shift research findings from the laboratory to the field. (Earlier versions of this chapter have been presented by Snijkers, 1996b, 1998.)

5. Computer-Assisted Qualitative Interviewing (CAQI), a method for pre-testing computerised questionnaires, is discussed in chapter 5. With CAQI a pre-test protocol is integrated in a computerised questionnaire to be tested. A CAQI protocol is expressed by instruction screens and probes built around the questions that are to be tested. CAQI creates realistic fieldwork conditions in the laboratory, and helps to conduct cognitive interviews in a standardised way. Our experience is that CAQI is a workable method, resulting in improved qualitative information on the question-and-answer process. (This chapter is a reprint of Snijkers, 1997a.)

In the next four chapters, case studies of cognitive research will be presented. Here, I will discuss the design and the results (i.e. identified problems in the questionnaire and recommendations for improvement) of these studies. Chapters 6 and 7 are reprints of reports presented to the client, showing how the research and its results have been presented.

6. In the chapters 6 and 7, laboratory studies using CAQI in-depth interviews will be described. In chapter 6 a questionnaire on income is pre-tested, using thinking aloud with follow-up probes, meaning-oriented probing and targeted test questions. (See Snijkers, 1994b.)
7. Chapter 7 describes a pre-test study on questions on daily activity, pension schemes, and education and training from the European Community Household Panel. In this study thinking aloud with follow-up probes, meaning-oriented probing and vignettes have been used. (See Snijkers, 1995c; Van de Donk & Snijkers, 1995.)
8. In chapter 8, a CAQI field study on the Continuous Survey on Living Conditions will be described. In this study all steps of the 5-step (pre-)test model of data collection development have been followed to develop and test the newly designed questionnaire and the data collection procedures. In this chapter I will concentrate on step 3, the qualitative operational field test, in which respondent reaction coding and targeted test questions have been used. In the preceding laboratory study (in step 2; Dehue, 1996; Akkerboom, Dehue & Snijkers, 1998) in-depth interviews and focus groups had been conducted. (This chapter is a reprint of Snijkers et al., 1999b, an extended version of Snijkers, 1996c.)
9. In chapter 9 an example of a focus group study will be presented. With this study a redesigned form of a business survey has been pre-tested: the Annual Establishment Production Survey. While in the other studies interviewer-administered questionnaires for persons and households were tested, here we have an example of applying cognitive methods to a business survey with a self-completion form. In this chapter, this study is presented within a discussion on a strategy of active respondent communication. This strategy is aimed at stimulating and motivating sampled respondent to participate in surveys. The role of a

cognitive laboratory in this strategy is to develop well-designed questionnaires, that look attractive and are easy to understand and complete, thus reducing response burden. Here, cognitive laboratory research will be placed within a broader view on Total/Tailored Survey Design and Total Quality Management. (This chapter is a reprint of Snijkers & Luppes, 2000.)

Chapter 10 concludes this thesis.

10. This chapter provides a summary of this thesis, including a summary of methods used and results of the pre-test studies. In this summary the identified problems are related to design errors in the questions, according to question design principles. Although pre-testing is a way to evaluate questionnaires and control for measurement errors, in the practice of survey design, its results – including recommendations for improving the questionnaire – are not always accepted. In this chapter, I will also discuss arguments against pre-test research, as well as strategies for presentation and gaining acceptance of this kind of research and its results, one of them being the application of pre-test methods according to scientific principles. To complete this thesis, I will discuss the improvement of the computerisation of questionnaires by describing a limited number of tests addressing the programming process of computer-assisted survey instruments and the human-machine interaction. Thus, 'validating' the questionnaire with respect to these issues. Finally, I will conclude this thesis by discussing ideas for future research, and by showing that pre-testing in a cognitive laboratory validates questionnaires.

References

Akkerboom, H., 1992, Proposals for a Questionnaire Lab for personal and household surveys. (In Dutch: Voorstellen voor een vragenlab voor persoons- en gezinsenquêtes. Internal CBS report, BPA no. 2155-92-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Akkerboom, H., 1996, Laboratory for the development and testing of Measuring Instruments. (In German: Labor für die Entwicklung und den Test von Erhebungsinstrumenten. In: Pretest und Weiterentwicklung von Fragebogen, Band 9 der Schriftenreihe Spektrum Bundesstatistik, Statistisches Bundesamt, pp. 66-71, Wiesbaden, Germany.)

Akkerboom, H., and F. Dehue, 1997, The Dutch Model of Data Collection Development for Official Surveys. (International Journal of Public Opinion Research, vol. 9, No. 2, pp. 126-145.)

Akkerboom, H., F. Dehue, and G. Snijkers, 1998, Qualitative Tests for Data Collection (Re-) Design. The Case of POLS. (In: Bakker & Winkels (eds.), Integration of Household Surveys: Design, Advantage and Methods, pp. 20-25.)

Akkerboom, H., and A. Luiten, 1996, Selecting Pre-testing Tools according to a Model of Questionnaire Development, with Illustrations concerning Patient Satisfaction with medical care. (In: Proceedings of the Conference of the American Association for Public Opinion Research, AAPOR, Salt Lake City.)

Bakker, B., and J. Winkels (eds.), 1998, Integration of Household Surveys: Design, Advantage and Methods. (Netherlands Official Statistics, Special Issue, Vol. 13, Summer 1998. Statistics Netherlands, Voorburg/Heerlen.)

Banks, R., C. Christie, J. Currall, J. Francis, P. Harris, B. Lee, J. Martin, C. Payne, and A. Westlake (eds.), 1999, … Leading Survey and Statistical Computing into the New Millennium. Proceedings of the Third ASC International Conference. (Association for Survey Computing, Chesham, Bucks, UK.)

Belson, W.A., 1981, The Design and Understanding of Survey Questions. (Gower Publishing Company, London.)

Belson, W.A., 1986, Validity in Survey Research. (Gower Publishing Company, London.)

Biemer, P., R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds.), 1991, Measurement Errors in Surveys. (Wiley, New York.)

Billiet, J.B., 1991, Social Cognition and the Requirement of a Pragmatic Theory of Question Wording, Form and Context Effects in Social Surveys (In: ’t Hart & Kox, Cognition and Survey Methodology, pp. 15-35.)

Blyth, B. (ed.), 1998, Market Research and Information Technology. Application and Innovation. (Esomar Monograph, No. 6, Amsterdam.)

Clark, H., and M. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), Questions about Questions, pp. 15-48.)

Converse, J.M., and S. Presser, 1986, Survey Questions. Handcrafting the Standardized Questionnaire. (Quantitative Applications in the Social Sciences Series, No. 63, Sage, Beverly Hills.)

Cox, B., D. Binder, B. Chinnappa, A. Christianson, M. Colledge, and Ph. Kott (eds.), 1995, Business Survey Methods. (Wiley, New York.)

De Heer, W. (ed.), 1996, Continuous Survey on Living Conditions: Qualitative Test Program. (In Dutch: POLS: Kwalitatief testtraject. Internal CBS report, BPA No. H4266-96-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Dehue, F., 1996, Cognitive Laboratory Test (In Dutch: Vragenlabtest. In: De Heer (ed.), Continuous Survey on Living Conditions: Qualitative Test Program, pp. 16-26.)

Dehue, F., 1997, Standard Protocols for Developing and Testing Data Collection Procedures. (In Dutch: Standaard protocollen voor de ontwikkeling en de test van dataverzamelings-procedures. CBS report, BPA No. H7247-97-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

De Leeuw, E., J. Hox, and G. Snijkers, 1995, The Effect of Computer-Assisted Interviewing on Data Quality. A Review. (Journal of the Market Research Society, Vol. 37, No. 4, pp. 325-344.)

De Leeuw, E., J. Hox, and G. Snijkers, 1998, The Effect of Computer-Assisted Interviewing on Data Quality. A Review. (In: Blyth (ed.), Market Research and Information Technology: Application and Innovation, pp. 173-198.)

Dippo, C., and J. Norwood, 1992, A Review of Research at the bureau of Labor Statistics. (In: Tanur, Questions about Questions, pp. 271-290.)

Ericsson, K.A., and H.A. Simon, 1984, Protocol Analysis: Verbal Reports as Data. (MIT press, Cambridge, Massachusetts.)

Forsyth, B.H., and J.T. Lessler, 1991a, A Taxonomy of Cognitive Laboratory Methods. (In: 't Hart & Kox, Cognition and Survey Methodolgy, pp. 64-85.)

Forsyth, B.H., and J.T. Lessler, 1991b, Cognitive Laboratory Methods: A Taxonomy. (In: Biemer et al. (eds.), Measurement Errors in Surveys, pp. 393-418.)

Fowler, F.J., and T.W. Mangione, 1990, Standardized Survey Interviewing. Minimizing Interviewer-related error. (Applied Social Science Research Methods Series, Vol. 18, Sage, Newbury Park.)

Gower, A., 1991, The Questionnaire Design Resource Centre's Role in Questionnaire Research and development at Statistics Canada. (Paper presented at the 48th Session of the International Statistical Institute, Caïro, September 1991.)

Groves, R.M., 1989, Survey Errors and Survey Costs. (Wiley, New York.)

Hippler, H.J., N. Schwarz, and S. Sudman (eds.), 1987, Social Information Processing and Survey Methodology. (Springer-Verlag, New York.)

Huigen, R., J. Nollen-Dijcks, and H. Akkerboom, 1994, Focus groups for qualitative research of the form for the Production Statistics of the Industry 1993. (In Dutch: Focusgroepen voor kwalitatief onderzoek van het vraagformulier Produktiestatistiek Industrie 1993. Internal CBS report, BPA no. 7657-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Jabine, T.B., M.L. Straf, J.M. Tanur, and R. Tourangeau (eds.), 1984, Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines. Report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology. (National Academy Press, Washington, D.C.)

Jobe, J.B., and D.J. Mingay, 1991, Cognition and Survey Measurement: History and Overview. (Applied Cognitive Psychology, Special Issue on Cognition and Survey Measurement, Vol. 5, No. 3, pp. 175-192.)

Luppes, M., 1993, Report of the visit to the Statistisches Bundesambt, visit to ZUMA Mannheim and participation in the workshop'Einführung in die Computerunterstützte Inhaltsanalyse'. (In Dutch: Verslag bezoek Statistisches Bundesambt, bezoek ZUMA Mannheim en deelname workshop 'Einführung in die Computerunterstützte Inhaltsanalyse'. Internal CBS report, BPA no. 13342-93-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Luppes, M., 1994, Interpretation and Evaluation of Advance Letters. (Paper presented at the 5th Workshop on Household Survey Non-response, Ottawa, Canada.)

Luppes, M., J. Nollen-Dijcks, H. Akkerboom, and G. Snijkers, 1994, Training in In-depth Interviewing Techniques. (In Dutch: Training in diepte-interviewtechnieken. Internal CBS report, BPA no. 3868-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), 1997, Survey Measurement and Process Quality. (Wiley, New York.)

Nollen-Dijcks, 1993, Voluntary Respondents and the Organisation of In-depth Interviews. (In Dutch: Proefrespondenten en de organisatie van diepte-interviews. Internal CBS report, BPA no. 12971-93-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Nollen-Dijcks, 1994a, Report of a Working Visit to Washington D.C., 20-24 September 1993. (In Dutch: Verslag van een werkbezoek aan Washington D.C., 20-24 september 1993. Internal CBS report, BPA no. 115-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Nollen-Dijcks, J., 1994b, 'Do we feel threatened?' In-depth Interviews with Questions from the Survey on Safety. Analysis Results. (In Dutch: 'Voelen we ons bedreigd?' Diepte-interviews met vragen uit de ERV; analyse-resultaten. Internal CBS report, BPA no. 4601-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Nollen-Dijcks, J., 1995, Focus Group Interviews as Qualitative Research Method. (In Dutch: Focusgroep-interviews als kwalitatieve onderzoeksmethode. CBS report, BPA no. H1904-95-DSH. Statistics Netherlands, Heerlen.)

Nollen-Dijcks, J., and G. Snijkers, 1994, 'Do we feel threatened?' In-depth Interviews with Questions from the Survey on Safety. Research Design. (In Dutch: 'Voelen we ons bedreigd?' Diepte-interviews met vragen uit de ERV; de onderzoeksopzet. Internal CBS report, BPA no. 113-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Schwarz, N., and S. Sudman (eds.), 1992, Context Effects in Social and Psychological Research. (Springer-Verlag, New York.)

Schwarz, N., and S. Sudman (eds.), 1994, Autobiographical Memory and the Validity of Retrospective Reports. (Springer-Verlag, New York.)

Schwarz, N., and S. Sudman (eds.), 1996, Answering Questions. Methodology for Determining Cognitive and communicative Processes in Survey Research. (Jossey-Bass, San Francisco.)

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds.), 1999, Cognition and Survey Research. (Wiley, New York.)

Sirken, M.G., T. Jabine, G. Willis, E. Martin, and C. Tucker (eds.), 1999, A New Agenda for Interdisciplinary Survey Research Methods: Proceedings of the CASM II Seminar. (National Center for Health Statistics, Washington, D.C.)

Snijkers, G. 1992, Computer-Assisted Interviewing: Telephone or Personal ? A Literature Study. (In: Westlake et al. (eds.), Survey and Statistical Computing, pp. 137-146.)

Snijkers, G., 1994a, 'What do we ask the voter?' Probing into Questions on the Election Survey. Analysis Results. (In Dutch: 'Wat vragen we de kiezer?' Doorvragen naar vragen uit het kiezersonderzoek. Analyse-resultaten. Internal CBS report, BPA no. 4139-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Snijkers, G., 1994b, What is the total net household income? Cognitive interviews on income questions. Design and Results. (In Dutch: 'Wat is het totale netto huishoudinkomen? Vragenlab-onderzoek naar de gecoördineerde verkorte inkomensvraag. Opzet en resultaten. Internal, unpublished CBS report. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1995a, What is the Total Net Household Income? Cognitive Interviews on Income Questions. (Presentation at the International Conference on Survey Measurement and Process Quality, Bristol, UK.)

Snijkers, G., 1995b, Pre-tests on ECHP Questions on Daily Activities, Pension Schemes, and Education and Training: Preliminary Results. (Internal CBS report, BPA no. H2117-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1995c, Pre-tests on ECHP Questions on Daily Activities, Pension Schemes, and Education and Training: Final Results. (Internal CBS report, BPA no. H8170-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1996a, The use of Computer-Assisted Interviewing in Cognitive Pre-tests of ECHP Questions. (Presentation at the Fourth International Social Science Methodology Conference, Colchester, Essex, UK.)

Snijkers, G., 1996b, Cognitively Pre-testing Questionnaires (In Dutch: Cognitief Pretesten van Vragenlijsten. Presentation at the NOSMO Methodologendag, Universiteit van Utrecht.)

Snijkers, G., 1996c, Report on Qualitative Field Test. (In Dutch: Verslag kwalitatieve veldproef. In: De Heer (ed.), Continuous Survey on Living Conditions: Qualitative Test Program, pp. 29-35.)

Snijkers, G., 1997a, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. (Bulletin de Methodologie Sociologique, No. 55, pp. 93-107.)

Snijkers, G., 1997b, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Testing of Computerised Questionnaires. (Presentation at the US Bureau of the Census, Washington DC, USA, and Statistics Canada, Ottawa, Canada.)

Snijkers, G., 1998, Cognitive Pre-testing of Questionnaires. (In Dutch: Cognitief pretesten van vragenlijsten. (Contribution to: Saris, W.E., Data Collection with Questionnaires. IOPS course, University of Amsterdam.)

Snijkers, G., H. Akkerboom, I. Kuijpers, and E. De Leeuw, 1996, Computer-Assisted Qualitative Interviewing: An "Intermediate" Technology of Quality Assessment. (Presentation at the The International Conference on Computer-Assisted Survey Information Collection (CASIC'96), San Antonio, Texas, USA.)

Snijkers, G., H. Akkerboom, and J. Nollen-Dijcks, 1994, 'What do we ask the Voter?' Probing into Questions on the Election Survey. Research Design. (In dutch: 'Wat vragen we de kiezer?' Doorvragen naar vragen uit het kiezersonderzoek. De onderzoeksopzet. Internal CBS report, BPA no. 114-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Snijkers, G., E. de Leeuw, D. Hoezen, and I. Kuijpers, 1999a, Computer-Assisted Qualitative Interviewing: Testing and Quality Assessment of CAPI and CATI Questionnaires in the Field. (Presentation at the Third International 3-Day Conference: … Leading Survey and Statistical Computing into the New Millennium. The Association for Survey Computing (ASC), University of Edinburg, Edinburgh, UK.)

Snijkers, G., E. de Leeuw, D. Hoezen, and I. Kuijpers, 1999b, Computer-Assisted Qualitative Interviewing: Testing and Quality Assessment of CAPI and CATI Questionnaires in the Field. (In: Banks et al. (eds.), … Leading Survey and Statistical Computing into the New Millennium, pp. 231-258.)

Snijkers, G., and M. Luppes, 2000, The Best of Two Worlds: Total Design Method and New Kontiv Design. An operational Model to improve Respondent Co-operation. (Paper presented at the Second International Conference on Establishment Surveys (ICES-II): Survey Methods for Businesses, Farms, and Institutions, June 17-21, Buffalo, New York.) (In: American Statistical Association, Proceedings of Invited Paper, pp. 361-371. Alexandria, Virginia.)

Sudman, S., N.M. Bradburn, and N. Schwarz, 1996, Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. (Jossey-Bass, San Francisco.)

Tanur, J.M. (ed.), 1992, Questions about Questions: Inquiries into the Cognitive Bases of Surveys. (Russel Sage Foundation, New York.)

Tanur, J.M., 1999, Looking Backwards and Forwards at the CASM Movement. (In: Sirken, Hermann et al. (eds.), Cognition and Survey Research, pp. 13-19.)

Tanur, J.M., and S.E. Fienberg, 1992, Cognitive Aspects of Surveys: Yesterday, Today and Tomorrow. (Journal of Official Statistics, Vol. 8, No. 1, pp. 5-17.)

't Hart, H., and W. Kox (eds.), 1991, Cognition and Survey Methodology. (Proceedings of the International Workshop, Utrecht University, November 18-19, 1991. Method Series MS-92-3. Utrecht University, Faculty for Social Sciences / ISOR, Utrecht, The Netherlands.)

Tourangeau, R., and K.A. Rasinksi, 1988, Cognitive Processes underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol. 103, No. 3, pp. 299-314.)

Van de Donk, P., and G. Snijkers, 1995, Testing on ECHP-questions on daily activities, pension schemes, and education and training: final report. (CBS report, BPA no. H2876-95-SIP. Statistics Netherlands, Division of Socio-Economic Statistics, Department of Integration and Presentation, Heerlen.)

Westlake, A., R. Banks, C. Payne, and T. Orchard (eds.), 1992, Survey and Statistical Computing. (North-Holland, Elsevier Science Publishers, Amsterdam.)

Willis, G.B., Th.J. DeMaio, and B. Harris-Kojetin, 1999, Is the Bandwagon headed to the Methodological Promised Land? Evaluating the Validity of Cognitive interviewing Techniques. (In: Sirken, Herrmann et al. (eds.), Cognition and Survey Research, pp. 133-153.)

Chapter 2
Aspects of Computer-Assisted Interviewing

Computer-Assisted Interviewing: Telephone or Personal?
A Literature Study

Summary: *This chapter discusses computer-assisted interviewing (CAI). The focus in this chapter is on comparing computer-assisted telephone and personal interviewing (CATI and CAPI), with regard to characteristics of these modes, costs and data quality. For these aspects the empirical research literature is reviewed.*
The discussion also includes general characteristics of CAI, like automated routing, complex skipping patterns and branching, tailoring of questions and question wording, range and consistency checks, calculations on answers and imputations, the possibility of last minute changes, and greater standardisation of the interview. Since at Statistics Netherlands most questionnaires are computerised, CAI sets the conditions for pre-test research at the Questionnaire Laboratory. This chapter provides a general overview of these conditions. Within this setting, we developed a method to pre-test computerised questionnaires: Computer-Assisted Qualitative Interviewing (CAQI), which will be discussed in chapter 5 and 8. CAI characteristics also hold for CAQI.

Keywords: *Computer-Assisted Interviewing (CAI), CAPI, CATI.*

- Paper presented at the First SGCSA (Study Group on Computers in Survey Analysis) International Conference on Survey and Statistical Computing, September 24-26, 1992, Bristol, UK.
- Paper published in: Westlake, A., R. Banks, C. Payne, and T. Orchard (eds.), 1992, Survey and Statistical Computing. The Conference Proceedings, pp. 137-146. (North-Holland, Elservier Science Publishers, Amsterdam.)
- The author is grateful to Abby Israëls, Randy Banks and two anonymous reviewers for their comments on earlier versions of this paper.

2.1. Introduction

Statistics Netherlands (CBS) uses three different interview modes to carry out its personal and household interviews: face-to-face interviews, in which an interviewer visits the respondent at home; telephone interviews, with the respondent being interviewed by telephone; and self-administered or mail interviews, in which the respondent fills out the questionnaire on his own. Each type of interview can be conducted in the classical way with paper-and-pencil, or with the computer. The resulting interviewing methods are shown in table 2.1 (Sikkel, 1986; Saris, 1989).

Table 2.1. Interviewing methods

Mode	Interviewing technique			
	paper-and-pencil (PAI)		computer-assisted (CAI)	
face-to-face	PAPI	(paper-and-pencil personal interviewing)	CAPI	(computer-assisted personal interviewing)
telephone	PATI	(paper-and-pencil telephone interviewing)	CATI	(computer-assisted telephone interviewing)
self- administered	PASI/PAMI	(paper-and-pencil self-administered/mail interviewing)	CASI/CAMI	(computer-assisted self-administered/mail interviewing)

Many CBS surveys use the PAPI or CAPI method. However, these methods are rather expensive, due to travel costs of the interviewers. In the literature, telephone interviewing is considered less expensive, with immediate availability of the data (Saris, 1989; Nicholls & Groves, 1986). In this chapter we shall compare CATI to CAPI. This will be done by means of a literature study. The questions that will be answered in this chapter are: What are the advantages and disadvantages of CATI compared to CAPI, and in what situations can CATI be used effectively?

Since there are very few studies in which CATI is compared to CAPI (Lyberg & Kasprzyk, 1991), we will first describe in section 2.2 the consequences of computer-assisted interviewing for traditional paper-and-pencil interviewing. In section 2.3 a comparison will be made between telephone and face-to-face interviewing. In section 2.4 we will conclude with a comparison between CATI and CAPI.

2.2. CAI versus PAI

It is very difficult to get a picture of the differences between CATI and CAPI based on papers about interviewing. Most papers do not compare CATI and CAPI but CATI and PAPI. As a consequence, more attention is paid to the comparison between CAI and PAI. The differences between CAI and PAI are described in a number of papers at several levels of detail (Lyberg & Kasprzyk, 1991; Saris, 1989, 1990; De Leeuw, 1989b; Sikkel, 1988; Nicholls & Groves, 1986; Groves & Nicholls, 1986; House, 1985; Kerssemakers, 1985). The specific characteristics of CAI which emerge from these papers are as follows:

1) The (complex) routing structure is implemented in the interview program. The program provides a rigid control over question flows and recording of responses, forcing the interviewer through question sequences appropriate to the respondent and demanding data entry to all questions presented. This removes the burden of following the correct routing from the interviewer.
2) Answers can be edited in the field by internal checks (range checks, consistency checks).
3) Question wording can be personalised with variable question texts.
4) There is the possibility of randomisation of questions and response categories.
5) There is the possibility of using existing data in follow up surveys, such as panel surveys.
6) There is the possibility of automatically coding answers to open-ended questions, as well as performing calculations on answers in the field.
7) An existing questionnaire (i.e. the interview program) can easily and quickly be changed, e.g. in case of last minute bugs.
8) The data are stored directly in the computer, which results (together with data editing) in a considerable reduction of time for data preparation and data cleaning.

In general, these characteristics are said to represent advantages of CAI over PAI. The quality of the data collected by CAI is considered better than for PAI (Lyberg & Kasprzyk, 1991).

In discussions on computer-assisted interviewing these advantages are said to hold for every questionnaire program. However, as in ordinary programming, there is a problem of scale involved in questionnaire design for CAI. A routing structure can become so complex that questionnaire designers have no clear view of the structure anymore. Furthermore, a very complex routing will result in a data matrix with a lot of holes in it and make the data difficult to analyse. Further, too many consistency checks in a questionnaire can make the interview program operate too slowly during an interview, and lead to the interview being interrupted too often by the program. Testing of questionnaires can also be very difficult for large and complex questionnaires. Testing of interview programs is, however, very important (especially after correcting last minute bugs), to ensure that the program will operate correctly in every situation. As for the fieldwork, the automatic routing makes it less flexible: the questionnaire is divided into segments with each question shown on a separate screen. There is in consequence no overview anymore and responses are forced to fit the formats of the data fields. Furthermore, since eye-hand co-ordination is lost on a computer, recording errors can take place while typing in responses.

Most of these advantages only hold when the interview program is well designed. When this is not the case the advantages can turn into their opposite. In some papers (House, 1985; Groves & Nicholls, 1986; Nicholls & House, 1987; House & Nicholls, 1988; Saris, 1990) it is stated that a “good” questionnaire can be obtained when the interview program has been designed and tested with the principles of modular structured programming, i.e. a top-down modular design. Tools for developing survey questionnaires are presented by Jabine (1985; flow charts) and Futterman (1988; graphs).

The design and testing of large questionnaires in computer-assisted interviewing is far more complex than for questionnaires in paper-and-pencil interviewing (Nicholls & Groves, 1986; Platek, 1985). The questionnaire designer has to write a computer program which defines every aspect of the interview in advance. This involves not only the routing structure, but also checks for incorrect answers and the means by which interviewers can correct answers in the field, (variable) question wording, screen layout, data entry field formats and data assignment to data fields. This means that the design phase for a questionnaire in CAl can last much longer than the design phase in PAI.

For a large survey organisation like Statistics Netherlands, extra investments in questionnaire design are not necessarily a disadvantage, e.g. in case of surveys with a repetitive character like continuing or panel surveys, and for standardised modules in questionnaires, like the household and education question boxes. Standardised questionnaire design offers a number of benefits, like improved cost-effectiveness and efficiency, comparability of data between different surveys, and simplification of interviewer training. These benefits, however, are dependent on the structure of the executive organisation. Extra investments in the design of questionnaire programs and standardisation are only worth while when there is a group of methodologists who do all the design work in a co-ordinated way rather than a situation in which every department/division invents the wheel over and over again. In addition to a central questionnaire design department, there has to be close collaboration between designers and subject matter experts on every individual survey, since it is necessary to have some knowledge of the target population, the target variables, and the topic of the survey in order to define the routing and the consistency checks.

2.3. Telephone versus face-to-face interviewing

Independent of whether or not it is computerised, the interview mode itself (face-to-face or telephone) has consequences for the design of the questionnaire, and costs and quality of the data. In subsection 2.3.1 we shall compare telephone to face-to-face interviewing. In subsections 2.3.2 and 2.3.3 costs and data quality aspects of these two interview modes are compared. An extensive discussion on telephone survey methodology is given in Groves et al. (1988).

2.3.1. Characteristics of telephone interviewing

1) The most important feature of telephone interviewing is the use of the telephone as interview medium, which means that only oral communication is possible. This has several limiting consequences.
   a. It is almost impossible to use visual aids, such as cards showing the response categories, during an interview. If it is absolutely necessary to use visual aids, materials can be mailed to the respondents before the telephone interview is conducted, although this is not recommended.

b. Since there is no visual contact between interviewer and respondent, breaks in a telephone interview while waiting for the next question or answer can be very annoying to both the respondent and the interviewer. This may also influence the quality of the data, as when the respondent tries not to slow the pace of the interview and takes too little time to come up with an answer. The interviewer may not take enough time to type in the complete answer to open-ended questions which may result in cryptic answers (Catlin, Ingram & Hunter, 1988; Körmendi & Noordhoek, 1989). Consequently, the pace of a telephone interview is in general higher than for face-to-face interviewing, which makes telephone interviewing only adequate for simple questions that can be answered instantaneously and which need little time to input the answers (i.e. closed-ended questions) (Nicholls & Groves, 1986).

2) CATI as compared to CAPI has several cost and quality advantages. Some of these advantages however only hold for centralised telephone interviewing, as is the case at Statistics Netherlands (Kerssemakers, 1985; Nicholls & Groves, 1986; Groves & Nicholls, 1986; De Leeuw, 1989a).
   a. The first advantage of CATl concerns the hardware equipment. A CATI system requires a computer network that can permanently be used for all surveys by all interviewers. For CAPI, by contrast, every interviewer has his or her own laptop computer which runs only one survey at a time. In CATI hardware and software are used more efficiently.
   b. The data are immediately available. There is no delay, and there are no costs caused by transmission of the data.
   c. There are no travel costs for the interviewers. This aspect is the one considered to save the most expenses (Groves, 1989) resulting in lower overall costs for CATI. Instead of travel costs, CATI generates telephone costs. These expenses will be higher for longer interviews.
   d. There are less interviewers required for the same number of interviews, since there is no time lost due to (unnecessary) travelling (when no one of the sample household is at home), and the interviews take less time. This results in a higher workload size for CATI interviewers.
   e. There are no problems in distributing interviewers over the country. Because of this, the interviewers can be randomly distributed; no regional effects occur and no cluster sampling is needed.
   f. Research on interviewer effects shows that the use of CATI reduces interviewer variance (Pannekoek, 1988; Stokes & Yeh, 1988; De Leeuw, 1989a), especially in centralised telephone interviewing (Groves, 1989).
   g. This last point can be explained by the enforcement of interviewing guidelines through constant monitoring and supervision of interviewers, which is another attractive feature of CATI.
   h. The same holds for online call scheduling and case assignment.

3) A third very important aspect of telephone interviewing concerns non-observation errors due to undercoverage (Massey, 1988; Trewin & Lee, 1988). Undercoverage emerges when households

belonging to the target population cannot be contacted due to gaps in the sample frame. Undercoverage occurs both for households who do not have a telephone and for those who have an unlisted number.

From CBS studies it became evident that the young, the elderly, and the unemployed do less frequently have a telephone (Kerssemakers et al., 1987). The same holds for divorced people, people with lower education levels, and people in lower income groups. Steel & Boal (1988) showed that in Australia low income households, younger people, and people living in rented houses are less accessible by telephone. In Britain (Wilson et al., 1988) telephone coverage is lower among the unemployed, even after controlling for age and social class. Groves (1989) concludes from the face-to-face U.S. National Health Interview Survey 1985-1986, that telephone coverage is lower among those aged under 25, with income under $15,000, non-whites, those who have never been married and those who are married but whose spouse is not in the household, and among those living in rural areas (due to higher monthly subscription costs in rural and remote areas).

4) Finally, using the telephone as communication medium makes some demands upon the questionnaire (Sudman & Bradburn, 1982; Kerssemakers, 1985; Converse & Presser, 1986; De Leeuw, 1989a; Körmendi & Noordhoek, 1989; Saris, 1990):
   a. Simple question wordings are recommended, as are short and clear instructions.
   b. Short lists of response categories in closed-ended questions, with a maximum of say 5 categories, are also recommended, since it is very difficult to recall a long list of categories on the phone. Long lists of categories will very likely result in picking the first, the last, or the first appropriate category. Numerical scales (e.g. 7-point scales in opinion questions) can be used in telephone surveys, although the direction of the scale might be altered by the respondent.
   c. The duration of an interview should not be too long. An average duration of 15-20 minutes, with upper limit 30 minutes, is indicated in the literature (Kerssemakers, 1985; De Leeuw, 1989a). De Heer et al. (1990) mention an average duration of telephone interviews of only 8 minutes, against up to 45 minutes in face-to-face interviewing. However, Körmendi & Noordhoek (1989) show that in Denmark a duration of 37 minutes did not gave any difficulties. Sudman & Bradburn (1982) state that telephone interviews may be of the same length as face-to-face interviews.

2.3.2. Costs

A systematic study comparing the costs of CATI and CAPI was not found, although there are comparisons between CATI and PAPI. Kerssemakers (1985) mentions reduction in costs of 40-50% for CATI compared to PAPI, when all aspects of the survey design (from survey and questionnaire design up to a clean data matrix) are incorporated. For interviews of short duration

the reduction in costs was a little higher (50-60%); for longer interviews cost-reduction was up to about 25%.

Groves (1989) concludes from a comparison of centralised PATI and PAPI from 1979 that telephone interviewing is about half as expensive as face-to-face interviewing. In this comparison only the costs of sampling and data collection are included, with the average length of the telephone interview about one-half of the face-to-face interview. The most important sources of cost savings in telephone interviewing are: no travel costs (about 20% of the total budget in the face-to-face survey), the higher workload size of interviewers (even when the length of interview is taken into account), less interviewers, and less administrative co-ordination of the data collection activities and materials, and general oversight of the staff. The only source in telephone interviewing generating extra costs, as compared to face-to-face interviewing, is the telephone communication costs (about 42% of the total budget in the telephone survey).

As for computer-assisted interviewing versus paper-and-pencil interviewing, Nicholls & Groves (1986) give figures on CATI and PATI. According to their findings, CATI neither reduces nor increases the costs. CATI generates higher start-up and running costs in hardware, software, training of staff and support. Cost savings can be found in interviewing and data processing. CATI may reduce costs for large, continuing surveys with simple questionnaires as compared to small, one-time surveys with complex questionnaires. According to Catlin & Ingram (1988) centralised CATI is slightly less expensive per household contacted than centralised PATI, and slightly more expensive per interview completed, although differences are small.

When CATI is related to CAPI under the same conditions (sampling design, number of respondents, and questionnaire), costs on software, support, design of the interview program, and data preparation and cleaning may be considered equal. CATI may be considered less expensive with respect to the number of interviewers, training and supervision, and data collection activities. On the other hand, telephone communication costs are higher for CATI: the longer the questionnaire, the higher these are these costs. Hardware costs are difficult to predict, but in centralised CATI hardware is more efficiently used than in CAPI.

2.3.3. Data quality

Quality aspects of interview data refer both to non-observation and observation errors. Under-coverage and non-response are non-observation errors. Observation errors are errors relating to interviewer and respondent behaviour, errors due the questionnaire, and mode effects. Observation errors are also referred to as measurement errors. Telephone interviewing undercoverage was already discussed in subsection 3.1.3. Non-response and measurement errors will be discussed below, concentrating on specific aspects of CATI. In comparison we will look at PAPI, since there is no material available on CAPI. The relation between non-response, undercoverage, and

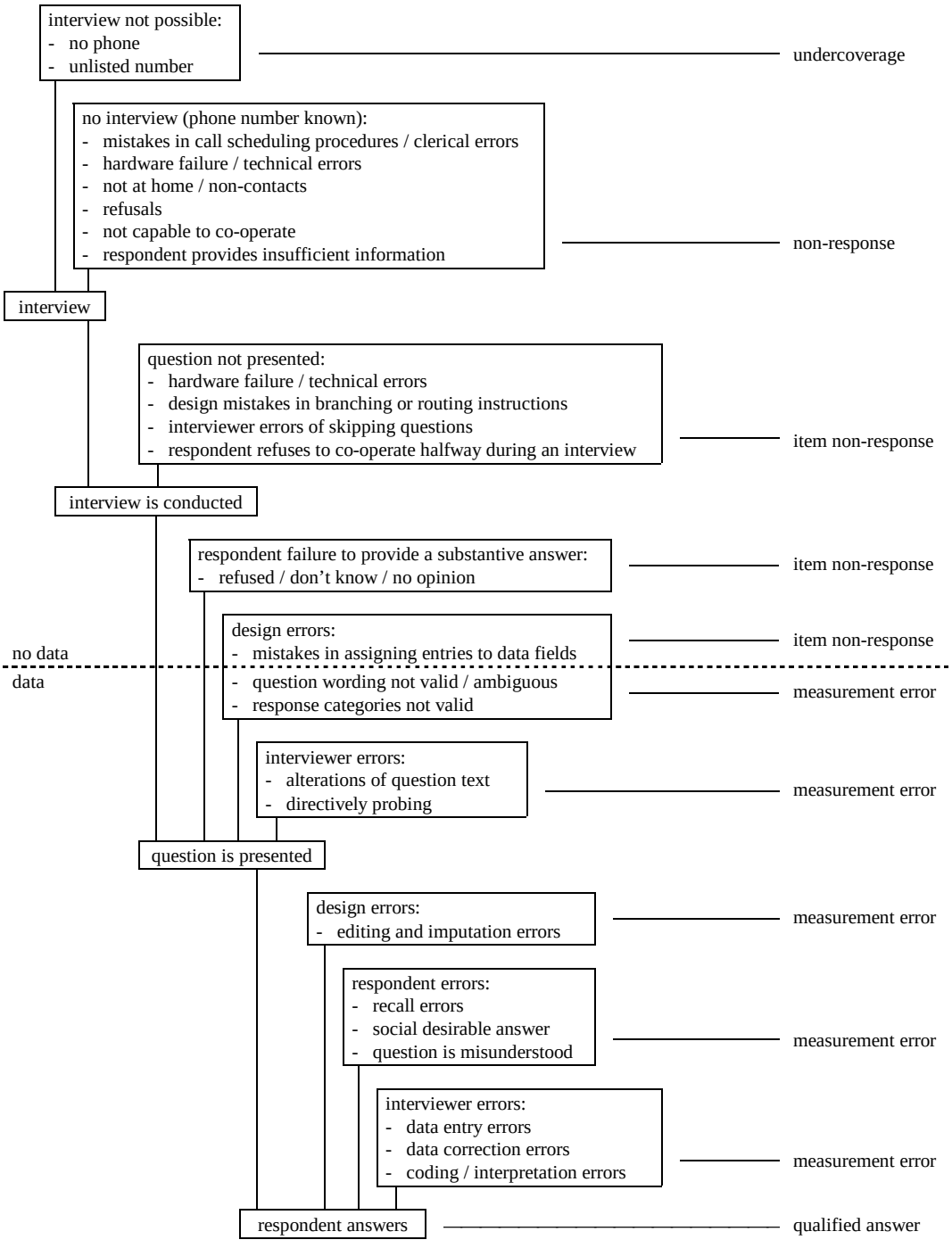
measurement error is given in figure 2.1. Extensive discussions on errors in surveys are given in Groves (1989) and Biemer et al. (1991).

Non-response in telephone surveys is affected by the same factors as in face-to-face surveys: the length of the questionnaire, the survey organisation and other design aspects. Apart from these factors, the telephone as communication medium may also affect response rates, especially refusal rates. This may be caused by age (with a higher proportion elderly persons refusing) and education (with lower response rates in lower education groups) (Groves & Lyberg, 1988). But De Leeuw (1990) concludes from PAPI and PATI experiments, that for important socio-demographic variables like gender, age, education and marital status, response rates do not differ between the two modes (i.e. no selective non-response). The response rates may also be influenced by the topic of the survey, especially when sensitive topics like income are covered. This may also result in item non-response for sensitive questions. De Leeuw (1990) reports no differences on item non-response between questions on income in PAPI and PATI. Kerssemakers (1985) and Körmendi & Noordhoek (1989) report higher item non-response figures for telephone surveys than for face-to-face interviewing. Van Eck & Kazemier (1990) show in their study on income from hidden labour that telephone interviewing is not appropriate for such a sensitive and complex topic.

Non-response figures in the literature on CATI are not consistent. Wilson et al. (1988) report higher non-response rates for CATI compared to PAPI. PAPI and PATI experiments from De Leeuw (1990) show higher response rates for PATI (66%) than for PAPI (51%). Groves & Lyberg (1988) mention several surveys with lower response rates for telephone surveys than for face-to-face surveys as well as surveys with higher response rates. According to Kerrsemakers (1985) response rates at Statistics Netherlands for both modes are the same. At the moment response rates at Statistics Netherlands for CATI (ranging from 55% to 75%) are even better than for CAPI (21% - 60%) (De Heer et al., 1990). Moreover, the overall non-response, refusal and non-contact rates for telephone surveys have not increased in the last decade, in contrast to rates for face-to-face surveys. Refusal and non-contact rates for telephone and face-to-face surveys are approximately the same (about 25% of the households refuses co-operation, and 7% of the households could not be contacted). Groves & Lyberg (1988), however, indicate that in telephone surveys refusals make up a higher proportion of the total non-response than in face-to-face surveys, since in telephone surveys non-contacts can be almost eliminated with repeated dialling. Most of these refusals take place in the very first minute of the conversation, after the introduction but before any explanation can be given of the purpose of the call.

Measurement errors arise from design errors in the questionnaire (as discussed in section 2.2 and subsection 2.3.1.4), interviewer and/or respondent behaviour, e.g. due to the use of the telephone as communication medium (see subsection 2.3.1.1). These aspects are shown in figure 2.1. According to Groves & Nicholls (1986), Körmendi & Noordhoek (1989), and Fowler & Mangione (1990) these effects can be countered by well designed interview programs (see section 2.2), considering the extra demands the telephone makes on the questionnaire (as discussed in

Figure 2.1. CATI observation and non-observation errors



subsection 2.3.1.4), well-trained interviewing staff, and good internal organisation with respect to online call scheduling and instructions, supervision and monitoring.

Several studies on mode effects on data quality show that the differences in data quality between telephone and face-to-face interviews are small. Körmendi & Noordhoek (1989) conclude that with good questionnaire design and interview training, “there appears to be no reason why the telephone interview cannot be a fully legitimate alternative to the face-to-face interview” (p. 85). From experiments on mode effects, De Leeuw (1990), and Sykes & Collins (1988) come to the same conclusion, as do De Leeuw & Van der Zouwen (1988) in a comparative meta-analysis on data quality in telephone and face-to-face interviewing.

In the past, telephone interviewing was considered to have some effect on socially desirable answers. De Leeuw & Van der Zouwen (1988) indicate that these differences have become smaller over time. In a newly performed meta-analysis, De Leeuw (1991) shows that in recent mode comparisons (after 1980) no differences were found for social desirability.

2.4. Summary and conclusions

The problem with a comparison of CATI related to CAPI based on literature is that are no studies available in which both interview methods are compared. We have tried to solve this problem by comparing computer-assisted interviewing (CAI) with paper-and-pencil interviewing (PAl) (in section 2.2), and telephone with face-to-face interviewing (in section 2.3). We have concluded that CATI will be cheaper than CAPI. Reduction in costs stem from lower personnel costs and the absence of travel costs. Telephone communication increases costs, which may be higher when interviews are longer. Data quality aspects seem to be mode independent (over the last decade). (item) non-response and measurement errors are dependent on the topic of the survey, Interviewer training and supervision, and the design of the interview program (to produce a “good” interview program is not an easy task to accomplish, as we have stated in section 2.2). The great advantages of centralised CATI are the reduction of Interviewer effects by supervision and monitoring of the interviews, the higher workload size of interviewers, and immediate availability of the data. Its weakness is that no visual aids can be used.

So, CATI appears to be a real alternative to CAPI, although the use of the telephone demands simple questions, with instant answering (no recall and lookup questions, no long lists of response categories), and limited data entry (few open-ended questions). Furthermore, CATI seems to be inappropriate when an average interview lasts longer than 30 minutes, and the target population of the survey has low telephone coverage.

References

Biemer, P.P., R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (eds.), 1991, Measurement Errors in Surveys. (Wiley, New York.)

Catlin, G., and S. Ingram, 1988, The Effects of CATI on Costs and Data Quality: A Comparison of CATI and Paper Methods in Centralized Interviewing. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 437-450.)

Catlin, G., S. Ingram and L Hunter, 1988, The Effect of CATI on Data Quality: A comparison of CATI and Paper Methods. (In: Proceedings of the Fourth Annual Research Conference, pp. 291-299. U.S. Bureau of the Census, Washington DC.)

Converse, J.M., and S. Presser, 1986, Survey Questions. Handcrafting the Standardized Questionnaire. (Quantitative Applications in the Social Sciences Series, No. 63, Sage, Beverly Hills.)

De Heer, W.F., J.C. Akkerboom and A.Z. Israëls, 1990, Ideas for Non-response Investigations. Contribution to the Non-response Workshop. (Internal CBS report. Statistics Netherlands, Heerlen/Voorburg).

De Leeuw, E.D., 1989a, A mail, telephone or face-to-face interview? A difficult choice. (In Dutch. In: Van der Zouwen, J., and W. Dijkstra (eds.), Social Science Research with Questionnaires. Methods, Bottlenecks and Solutions, pp. 17-34.)

De Leeuw, E.D., 1989b, Computer-Assisted Surveys: a review of the technological changes in data collection procedures. (In Dutch. Tijdschrift voor Onderwijsresearch, Vol. 14, No. 4, pp. 201-213.)

De Leeuw, E.D., 1990, Data Quality in Mail, Telephone and Face-to-Face Surveys; A Mode Comparison in the Netherlands. (Response Effects in Surveys, Technical Report No. 4, Free University, Amsterdam.)

De Leeuw, E.D., 1991, Data Quality of Mail, Telephone and Face to Face Surveys; A Quantitative Review. (Free University, Interuniversity Institute for Psychometrics and Sociometrics/ Department of Research Methods, Amsterdam.)

De Leeuw, E.D., and J. van der Zouwen, 1988, Data Quality in Telephone and Face-to-Face Surveys: A Comparative Meta-Analysis. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 283-299.)

Fowler, F.J., Jr., and Th. W. Mangione, 1990, Standardized Survey Interviewing. Minimizing Interviewer-Related Error. (Sage Publications, Newbury Park, California.)

Futterman, M., 1988, CATI Instrument Logical Structures: An Analysis with Applications. (Journal of Official Statistics, Vol. 4, No. 4, pp. 333-348.)

Groves, R.M., 1989, Survey Errors and Survey Costs. (Wiley, New York.)

Groves, R.M., P.P. Biemer, LE. Lyberg, J.T. Massey, W.L Nicholls and J. Waksberg (eds.), 1988, Telephone Survey Methodology. (Wiley, New York.)

Groves, R.M., and L. Lyberg, 1988, An Overview of Non-response Issues in Telephone Surveys. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 191-211.)

Groves, R.M., and W.L Nicholls, 1986, The Status of Computer-Assisted Telephone Interviewing: Part II - Data Quality Issues. (Journal of Official Statistics, Vol. 2, No. 2, pp. 117-134.)

House, C.C., 1985, Questionnaire Design with Computer Assisted Interviewing. (Journal of Official Statistics, Vol. 1, No. 2, pp. 209-219.)

House, C.C., and W.L Nicholls, 1988, Questionnaire Design for CATI: Design Objectives and Methods. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 421-436.)

Jabine, Th. B., 1985, Flow Charts - A Tool for Developing and Understanding Survey Questionnaires. (Journal of Official Statistics, Vol. 1, No. 2, pp. 189-207.)

Kerssemakers, F.A.M., 1985, Telephone interviewing. (In Dutch. In: CBS Select 3, pp. 211-230. (Statistics Netherlands, Vooburg/Heerlen.)

Kerssemakers, F.A.M., F.A.C. de Mast and P.W.M. Remmerswaal, 1987, Computer-Assisted Telephone Interviewing, some response findings. (In: CBS Select 4, pp. 119-131. Statistics Netherlands, Vooburg/Heerlen.)

Körmendi, E., and J. Noordhoek, 1989, Data Quality and Telephone Interviews. A comparison study of face-to-face and telephone data collection methods. (Danmarks Statistik, Copenhagen.)

Lyberg, L, and D. Kasprzyk, 1991, Data Collection Methods and Measurement Error: An Overview. (In: Biemer et al. (eds.), Measurement Errors in Surveys, pp. 237-257.)

Massey, J.T., 1988, An Overview of Telephone Coverage. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 3-8.)

Nicholls, W.L, and R.M. Groves, 1986, The Status of Computer-Assisted Telephone Interviewing: Part 1 - Introduction and Impact on Cost and Timeliness of Survey Data. (Journal of Official Statistics, Vol. 2, No. 2, pp. 93-115.)

Nicholls, W.L, and C.C. House, 1987, Designing Questionnaires for Computer-Assisted Interviewing: A Focus on Program Correctness. (In: Proceedings of the Third Annual Research Conference, pp. 95-111. U.S. Bureau of the Census, Washington.)

Pannekoek, J., 1988, Interviewer variance in a telephone survey. (Journal of Official Statistics, Vol. 4, No. 4, pp. 375-384.)

Platek, R., 1985, Some Important Issues in Questionnaire Development. (Journal of Official Statistics, Vol. 1, No. 2, pp. 119-136.)

Saris, W.E., 1989, A Technological Revolution in Data Collection. (Quality and Quantity, Vol. 23, No. 3-4, pp. 333-349.)

Saris, W.E., 1990, Microcomputers in Data Collection (first draft). (University of Amsterdam, Department of Statistics and Methodology, Amsterdam.)

Sikkel, D., 1986, Computer-Assisted Data Collection. (CBS report. Statistics Netherlands, Department of Statistical Methods, Voorburg.)

Sikkel, D., 1988, Quality Aspects of Statistical Data Collection. (Ph.D.-thesis. Sociometric Research Foundation, Amsterdam).

Steel, D., and P. Boal, 1988, Accessibility by Telephone in Australia: Implications for Telephone Surveys. (Journal of Official Statistics, Vol. 4, No. 4, pp. 285-297.)

Stokes, L, and M-Y. Yeh, 1988, Searching the Causes of Interviewer Effects in Telephone Surveys. (In: Groves et al. (eds.), Telephone, Survey Methodology, pp. 357-373.)

Sudman, S., and N.M. Bradburn, 1982, Asking Questions. A Practical Guide to Questionnaire Design. (Jossey-Bass Publishers, San Francisco.)

Sykes, W., and M. Collins, 1988, Effects of Mode of Interview: Experiments in the UK. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 301-320.)

Trewin, D., and G. Lee, 1988, International Comparisons of Telephone Coverage. (In: Groves, et al. (eds.), Telephone Survey Methodology, pp. 9-24.)

Van der Zouwen, J., and W. Dijkstra (eds.), 1989, Social Science Research with Questionnaires. Methods, Bottlenecks and Solutions. (In Dutch. VU Publishers, Amsterdam.)

Van Eck, R., and B. Kazemier, 1990, Income from Hidden Labour. An Analysis of Survey Results. (Statistical Essays M36, Statistics Netherlands, Voorburg/Heerlen.)

Wilson, P., N. Blackshaw and P. Norris, 1988, An Evaluation of Telephone Interviewing on the British Labour Force Survey. (Journal of Official Statistics, Vol. 4, No. 4, pp. 385-400.)

Chapter 3

Quality Aspects of Survey Data in Computer-Assisted Interviewing

The Effect of Computer-Assisted Interviewing on Data Quality

A Review

Edith de Leeuw, Joop Hox, and Ger Snijkers

Summary: *In addition to the last chapter this chapter also focuses on Computer-Assisted Interviewing. This chapter summarises what was known about computer-assisted data collection methods (CADAC) in 1995. Since we also use the computer to conduct cognitive interviews (in CAQI, see chapter 5), the emphasis is on the effect of the computer on survey data quality and acceptance of the computer by respondents and interviewers. The chapter starts with a taxonomy of the various computer-assisted data collection methods and a discussion on data quality. This is followed by a review of factors influencing survey data quality. Subsequently for each of the principal CADAC methods in use the empirical research literature is reviewed. The main conclusions are that computer-assisted data collection methods are accepted by both respondents and interviewers, and that survey data quality improves, especially when complex questionnaires are used. In general, respondents are positive about the use of the computer during an interview: they attribute a greater degree of professionalism to the interview. The social interaction with the interviewer is described as comfortable and relaxed. As for the interviewers, the computer makes additional interviewer training in computer usage and computer-assisted interviewing necessary.*

Keywords: *Computer-Assisted Data Collection, CADAC, CATI, CAPI, CASI, DBM, EMS, VR, ASR, Telepanel, Acceptance by Respondent, Acceptance by Interviewer, Response Rate, Data Quality, Costs.*

- Paper published in: Journal of the Market Research Society, 1995, Vol. 37, No. 4, pp. 325-344.
- Paper published in: Blyth, B. (ed.), 1998, Market Research and Information Technology. Application and Innovation. (ESOMAR, No. 6, pp. 173-198, Amsterdam.)
- The authors thank an anonymous reviewer for his comments; they acknowledge the kind and helpful assistance of ESOMAR in locating relevant studies for this review.

3.1. Introduction

Whether computer-assisted data collection methods should be used in social research is no longer an issue. Most professional research organisations, commercial, government and academic, are adopting these new methods with enthusiasm. Computer-assisted telephone interviewing (CATI) is most prevalent and computer-assisted personal interviewing (CAPI) is rapidly gaining in popularity. Also, new interesting forms of computerised data collection, for instance automatic speech recognition, are emerging. This raises the question, what influence computer-assisted data collection methods have on the quality of the data. In this article we review the literature on this topic, focusing mainly on the three basic forms of data collection in surveys: the face-to-face interview, the telephone interview, and the (mail) questionnaire. For a review of electronic observation techniques such as barcode scanning, or automatic registration of TV watching (people or TV-meter) we refer to Saris (1989). For an interesting discussion on the potentials of people meters that goes beyond mere registration and into full-fledged media research see Samuels (1994).

We start our review with a taxonomy of different types of computer-assisted interviewing and a discussion of data quality. Next, we present a model of the factors that may lead to differences in data quality between computer-assisted and traditional interview procedures. Subsequently, we give an overview of the results of empirical research on data quality differences. Finally, we discuss the consequences of our findings for social and market research.

3.2. Taxonomy of different forms of data collection

Computer-assisted methods for survey research are often summarised under the global terms CADAC (Computer-Assisted Data Collection), CASIC (Computer-Assisted Survey Information Collection), and CAI (Computer-Assisted Interviewing); in this context the traditional paper-and-pencil methods are often denoted by PAPI (Paper-And-Pencil Interviewing). For a comparative review, see Weeks (1992); reviews that also offer some insight in the usage of computer-assisted methods in European countries are *inter al* Hippler & Beckenbach (1992), Hox et al. (1990), and Porst et al. (1994). Characteristic for all forms of computer-assisted interviewing is that questions are read from the computer screen, and that responses are entered directly in the computer, either by an interviewer or by a respondent. An interactive program presents the questions in the proper order, which may be different for different (groups of) respondents. There are three survey modes where CADAC may be employed: telephone, personal or face-to-face, and self-administered. Table 3.1 presents a systematic overview of the various computer-assisted interviewing methods.

Table 3.1. Taxonomy of Computer-Assisted Interviewing Methods

Specific method	Computer-assisted method
face-to-face interview	CAPI (Computer-Assisted Personal Interviewing)
telephone interview	CATI (Computer-Assisted Telephone Interviewing)
written questionnaire (general self-interview)	CASI (Computer-Assisted Self Interviewing) Synonyms: CSAQ (Computer Self-Administered Questionnaire), PDE (Prepared Data Entry)
mail survey	DBM (Disk-by-Mail), EMS (Electronic Mail Survey)
panel research	CAPAR (Computer-Assisted Panel Research), Tele-interview, (Electronic diaries)
various (no interviewer)	TDE (Touchtone Data Entry), VR (Voice Recognition), ASR (Automatic Speech)

3.2.1. Computer-Assisted Telephone Interviewing (CATI)

Computer-Assisted Telephone Interviewing (CATI) is the oldest form of computer-assisted interviewing (cf. Nicholls & Groves, 1986). Originally CATI would be employed centrally using a minicomputer system. Each interviewer sits behind a terminal and asks the questions that appear on the screen; the respondent's answer is then typed into the computer by the interviewer.

Supervisors are present for quality control and to assist with specific problems. This is still the most usual CATI setup, with a microcomputer network replacing the minicomputer system. However, the technological change to personal microcomputers makes it also possible to conduct a decentralised CATI survey, for instance from the interviewers' own homes.

3.2.2. Computer-Assisted Personal Interviewing (CAPI)

In CAPI, interviewers visit respondents with a portable computer (generally a notebook) and conduct a face-to-face interview using the computer. After the interview the data are sent to a central computer, either electronically by modem or by sending a data disk by mail (cf. Baker, 1992; Saris, 1991).

3.2.3. Computer-Assisted Self Interviewing (CASI)

Computer-assisted Self Interviewing (CASI) is also called Computerised Self-Administered Questionnaires (CSAQ) or Prepared Data Entry (PDE). Characteristic for CASI is that the respondents themselves read the questions on the screen and enter the answers. There is no interviewer; the interviewing program guides the respondent through the questionnaire.

CASI can appear as part of a CAPI session where 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 procedure used in traditional PAPI face-to-face interviews where an interviewer might give the respondent a paper questionnaire containing sensitive questions.

Two different computer-assisted equivalents of the mail survey are the Disk-by-Mail (DBM) and the Electronic Mail Survey (EMS).[1] In DBM a disk containing the interviewing program is sent to the respondent, who runs the program on his or her own computer and then returns the disk with the responses (cf. Higgins et al., 1987).[2] It is obvious that at present this works only with special populations, who have access to a computer. These may be private persons, but also corporations. Especially within corporations this method is used to collect inventory data for recurring stock-taking (cf. Weeks, 1992). DBM has also been used in surveys of teachers, using the school's computer (cf. De Leeuw, 1989) and of experienced PC-users (Jacobs, 1993).[3]

In EMS the survey is sent by electronic mail through existing computer networks, electronic mailing systems, and bulletin boards. Users of such systems receive a request to participate in a survey, and if they comply, they either are asked a number of questions by an interviewing program, or they receive an electronic form to fill in at a later stage. This is at present only possible with special populations, but the limited experience so far is positive (cf. Kiesler & Sproull, 1986).[4]

A fourth form of CASI is the Tele-interview (Saris, 1991). This is a form of computer-assisted panel research (CAPAR) where respondents fill in an electronic questionnaire about once a week.[5] For this, a large number of selected households receive a microcomputer and a modem. At regular intervals, the modem automatically queries a remote computer, and the computer receives new questionnaires for selected members of the household. After the questionnaires have been answered using the interviewing program, the data are sent back to the remote computer. For questions and technical problems a help desk is available through a toll-free number. The tele-interview has the advantage that it is not confined to special populations with access to computers. However, the tele-

[1] The latest form of CASI is Computer-Assisted Web Interviewing (CAWI), in which questionnaires are dispersed via the Internet (see e.g. Dillman, 2000).
[2] In 2000 at Statistics Netherlands, Snijkers et al. (unpublished) carried out a small-scale field study in which 50 selected households, who had participated in the 1999 Expenditure Survey, received a CD-rom by mail, containing the questionnaire program (in Blaise). The questionnaire could automatically be installed on the respondent's computer. After the questionnaire was completed, the data had to be returned to Statistics Netherlands (CBS) via the Internet by logging-in on a CBS server.
[3] The DBM survey of Jacobs was a survey of experienced PC-users on the amount of illegally copied software they owned. Thus, DBM was not only practical with this population, but also profitable because it ensured complete privacy.
[4] In 1998 and 1999, Snijkers et al. carried out two EMS studies with establishment surveys. In the 1998 study (see Roos, Jaspers & Snijkers, 1999; Wings & Snijkers, 1998), the monthly Survey on Short Term Economical Indicators was sent to businesses by e-mail. The respondents had to fill-in the data and return the completed questionnaire by e-mail. In the 1999 study (see Roos & Snijkers, 1999; Wings & Snijkers, 1998) construction enterprises received an e-mail containing a password and an ID, making it possible to log-in on a CBS server via the Internet in order to reach the monthly Survey on Construction Businesses. This Blaise questionnaire could be completed on-line or off-line.
[5] In CAPAR the tele-interview may be alternated by CAPI or CATI, thus leading to a mixed-mode design.

interview shares all methodological problems of traditional panel research (see Kaspzryk et al., 1989), although experience has shown that the bonus of having a free home computer leads to very low panel loss. The Dutch NIPO research institute operates a highly successful 'Telepanel' since 1987 (cf. Van Doorn, 1987; Van Doorn & Hess, 1989). The 'Gallupchannel' in Finland is another example (cf. Samuels, 1994). A variation on the tele-interview is an electronic diary for time budget and consumer behaviour research (Kalfs, 1993).

Both Weeks (1992) and Saris (1991) mention two very specific applications of CASI: Touchtone Data Entry (TDE) and Voice Recognition (VR) or Automatic Speech Recognition (ASR). In the first case a respondent is called by a computer, the questions are asked by a computer voice, and the responses are given by punching the appropriate number. In VC the respondent has to answer 'yes' or 'no' verbally. Automatic Speech Recognition has far more potential; in ASR a large vocabulary of meaningful words, such as holiday destinations, can be understood and acted upon by the interview system (Blyth & Piper, 1994). For a review and comparison with interviewer techniques, see Havice & Banks (1991).

3.3. Quality criteria

There can be no dispute that we should aim for high quality data. But what exactly defines high quality? Can there be any absolute standards? Deming, in 1944, already stressed that absolute 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.

A review of the literature on survey error and data quality (Groves, 1989) reveals four identifiable sources of error: coverage, non-response, sampling and measurement or response error. The effects of measurement error and to a lesser degree non-response error can far exceed those of sampling error (cf Denny & Galvin, 1993) and can be a serious threat for data quality. When studying factual behaviour, it is sometimes possible to estimate the precise accuracy of the data by comparing the survey results with some record or observational data. Usually such data are not available and for research on attitudes, opinions and intentions this criterion for data quality is impossible to use. In those cases other criteria have been used such as item-missing data, detail on open questions, and social desirability. Groves (1989) also explicitly incorporates cost considerations into the discussion of survey quality.

When writing a review one is always restricted by the content and the amount of detail in the original articles. Until now, research articles on computer-assisted interviewing have mainly focussed on the acceptability of this new technology and the consequences for response rates and on item-missing data. In a few cases social desirability and detail on open questions have been investigated too. Systematic cost-estimates are rare. Therefore, this review is limited to data quality as indicated by

high acceptability, low unit and item non-response and high detail on open questions (completeness) and low social desirability bias. Whenever data were available costs have also been considered.

3.4. A model for the influence of CADAC on data quality

Computer-assisted interviewing has become rapidly popular partly because of the expectation that it would lead to better data quality than traditional methods (cf. Denny & Galvin, 1993). A priori there are three groups of factors that may affect data quality: (1) the technological possibilities of CADAC programs, (2) the visible presence of a computer, and (3) the effect of CADAC on the interviewing situation. The consequences of CADAC for costs will also be addressed.

3.4.1. Technological possibilities

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 do 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 CADAC 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' to a seven-category Likert scale, 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 analyse 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 PAPI, internal validity checks have to be conducted in 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. In a CADAC session there is an opportunity to correct range and consistency errors, and therefore CADAC should lead to fewer data entry errors and missing data.

3. The computer offers new possibilities for formulating questions. One example is the possibility to randomise 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 randomised, which avoids question format effects (e.g. recency 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 are awkward to use in PAPI methods.

4. There is no separate data entry phase. This means that the first tabled results can be available soon after the data collection phase. On the other hand, construction and programming of the questionnaire takes considerable time in CADAC. Thus, a well-planned CADAC survey has a real advantage when the results must be quickly available (as in election forecasts).

5. 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 from 'cheating'. Computer-assisted interviewing provides a research organisation with greater interviewer control and offers some protection against undesirable interviewer behaviour.

3.4.2. Visible presence of the computer

The visible presence of a computer may affect the data quality, apart from the technical aspects of using a computer. As with most technological innovations these effects are temporary. After some time everybody gets used to the new machine, and its influence on the situation is small. Clearly we are currently in a transition period; the computer is no longer an unimaginable technological wonder, but it is also not yet a common household item.

Compared to the traditional PAPI methods, the visible presence of a computer could lead to four effects on the way the respondents or the interviewers perceive the interview situation.

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 CAPI, there was considerable concern about this effect.

2. More privacy. Using a computer could also lead to expectancy of greater privacy by respondents; responses are typed directly into the computer and cannot be read by anyone who happens to find the questionnaire. In the western world, where computers are widespread and familiar, this reaction is more likely than the 'big brother' reaction.

3. Trained interviewers may feel more self-confident using a computer, and behave more professionally. This could lead to more confidence of the respondent in the interviewing

procedure. Social exchange theory, as applied to the survey process (Dillman, 1978) predicts that this should lead to more willingness to comply with the interviewers' requests.

4. In panel research the availability of a free home computer acts as a reinforcement for the respondents to continue to participate faithfully. Disk-by-mail (DBM) and Electronic Mail Surveys (EMS) both have a strong novelty effect: the survey request is highly visible, and not likely to be incorrectly perceived as junk mail. This should lead to a higher willingness to participate.

3.4.3. Effect of the computer on the interview situation

The effect of CADAC on the interview process depends strongly on the amount of training and/or experience the interviewers have with this method of data collection.

Inexperienced interviewers may direct much of their attention to keeping the computer running and correctly typing in the answers. If interviewers cannot touch-type, typing in long answers may lead to less eye contact between interviewers and respondents, causing the interviewers to miss non-verbal reactions of the respondents. If the computer is located between the interviewer and the respondent, even the physical distance may be greater than in PAPI. The methodological survey literature stresses the importance of good (non-verbal) communication and rapport between interviewers and respondents. If using the computer weakens the relation between interviewer and respondent, the interview will not be conducted optimally, and in consequence the data quality may suffer.

On the other hand, an experienced interviewer can rely on the computer for routings and complex question sequences, and therefore pay more attention to the respondent and the social processes involved in interviewing. Sometimes, for instance in asking sensitive questions, less eye-contact is an advantage (cf. Argyle & Dean, 1965); experienced interviewers can use the presence of a computer to their advantage by directing their attention to the screen when asking sensitive questions.

The conclusion is that in a CADAC survey we need interviewers who are well trained and experienced in computer-assisted data collection techniques. This means that in addition to a thorough basic interview training, an additional training in computer usage and computer-assisted interviewing is needed (see Woijcik et al., 1992, for a training program for computer-assisted interviewing). Given well-trained and experienced interviewers, the altered interview situation is likely to have more advantages than disadvantages, especially with sensitive questions.

3.4.4. Time and money: CADAC and its consequences for costs

Going from paper-and-pencil to computer-assisted interviewing requires initial investments, not only in equipment, but also in time. One has to invest in hardware, in software and in acquiring the hardware- and software-related knowledge and skills.

As mentioned above, basic interviewer training now needs to include training in handling a computer and using the interview software. But in contrast to this extended general interview training, training for actual surveys is less costly. Many topics (skipping, branching, selection rules) need not be taught because the interview software now handles this (see Porst et al., 1994). Executives, research directors and field managers also have to learn and appreciate computer-assisted interviewing.

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

In other words, developing, implementing and testing the questionnaire is more expensive, but no data entry is needed, and data editing and data cleaning cost less. 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. Samuels (1994) mentions a reduction of delivery time of 50% for the results of an omnibus survey. When time considerations and fast release of results are important for a client, this is an important advantage of CADAC over paper-and-pencil methods.

In the sections below, we will review the results of empirical comparative research on the effects of computer-assisted interviewing versus paper-and-pencil methods on data quality, in face-to-face, telephone, and self-administered interviews. Since acceptance of the computer-assisted methods is an important criterion in itself, we will also include research on the attitudes and opinions of interviewers and respondents. When possible, data on cost comparisons have been added.

3.5. Computer-Assisted Personal Interviewing (CAPI)

3.5.1. Effect on the respondent

Although the first users of CAPI were afraid of a negative effect on the response rate, even in the first applications of the method in Sweden and the Netherlands this did not occur (Van Bastelaer et al., 1987, p. 39). Later studies confirm that CAPI and paper-and-pencil methods yield comparable response rates in studies in the USA (Bradburn et al., 1992; Sperry et al., 1991; Thornberry et al., 1991), England (Martin et al., 1993), Sweden (Statistics Sweden, 1989) and Germany (Riede & Dorn, 1991). 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. Baker (1990, 1992) reports that most respondents find CAPI interesting and amusing, 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. When explicitly asked about the data privacy, 47% have more trust in the privacy of computer collected data, 5% have more trust in traditionally collected data, and 48% sees no difference.

Beckenbach (1992) conducted a small scale and well-controlled study comparing CAPI, CSAQ and a paper-and-pencil face-to-face interview. After the interview, both interviewers and respondents filled in a questionnaire with questions about the interview itself. Neither interviewers nor respondents report problems with eye contact or social interaction. In the computer-assisted methods (both CAPI as CSAQ) respondents were more positive about data privacy, and judged answering sensitive questions as less unpleasant.

3.5.2. Effect on the interviewer

Interviewers are in general markedly positive about computer-assisted interviewing. They appreciate the support that a good CAPI system offers when complex questionnaires are employed (Riede & Dorn, 1991; Edwards et al., 1993), they like working with the computer (Martin et al., 1993), and derive a feeling of professionalism from it (Edwards et al., 1993). Riede and Dorn (1991) point out that the one important complaint by interviewers is about the difficulty of grasping the overall structure of the questionnaire. CADAC questionnaires are typically screen-oriented, and it is not always possible to backtrack to earlier sections of the questionnaire for corrections or additions to earlier answers. Advanced CADAC programs have this flexibility, but they still have more constraints than paper-and-pencil methods (Weeks, 1992).

The studies in the previous paragraph all employed well-trained and computer-experienced interviewers. This is important, because Van Bastelaer et al. (1987) found clear differences between interviewers with and without experience in computer-assisted interviewing. They report that in the first week of data collection the percentage of interviewers that prefers CAPI was 52%, while in the third week this percentage had increased to 71%. An intensive training in using the computer and the specific CADAC program is essential (cf. Bennet & Goodger, 1993; Woijcik et al., 1992). With good training, older interviewers and interviewers without any previous computer experience can also enjoy using the computer and conduct good interviews (Edwards et al., 1993).

At first, interviewers may experience problems with open-ended questions. When they are not keyboard literate and lack typing skills, entering a detailed answer to an open-ended question can be slow and laborious. However, when interviewers gain keyboard experience they become fast enough typists to correctly record answers verbatim (Bond, 1991; Denny & Galvin, 1993).

The weight of the computer is sometimes mentioned as a problem (Edwards et al., 1993). In a study of the ergonomical aspects of microcomputers used in computer-assisted interviewing Couper and Groves (1992) also conclude that weight is an important ergonomic factor. Other ergonomic aspects have also been investigated. Beckenbach (1992) reports that 80% of the interviewers have no problems with screen or keyboard, while 75% report no problem at all. Finally, in the comparative study by Edwards et al. (1993) about three in four interviewers report that they found PAPI more tiring!

3.5.3. Effect on data quality

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). In addition, a well-implemented CAPI system prevents many interviewer mistakes. As a result, we may expect that, compared to traditional paper-and-pencil methods, computer-assisted interviewing have a positive effect on data quality.

Empirical studies tend to confirm this expectation. The percentage missing data is clearly lower in CAPI, mostly because interviewers cannot make routing errors (Sebestik et al., 1988; Olsen, 1992). In a pilot CAPI study, Bradburn et al. (1992) find that the number of missing data caused by respondents (‘don’t know’, ‘no answer’) also diminishes, but in the main study this is not replicated (Baker & Bradburn, 1992; Olsen, 1992). Other studies also fail to find a difference in respondent-induced missing data (Bemelmans-Spork et al., 1985; Martin et al., 1993).

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

An early comparative study by Waterton (1984, see also 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 presumably means that 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 more like a self-administered questionnaire. In the PAPI mode the question was asked and the answer taken down by the interviewers. Since self-administered questionnaires typically show less social desirability bias than face-to-face interviews (De Leeuw, 1993), the difference between PAPI and CAPI in this study may well correspond to a difference between an interview and a self-administered questionnaire. Other studies that compare PAPI and CAPI more precisely report slightly less social desirability bias with CAPI (Baker & Bradburn, 1992; Bradburn et al., 1992; Martin et al., 1993), but the differences are very small, generally smaller than differences typically found in comparisons of face-to-face versus telephone interviews or experienced versus inexperienced interviewers (Olsen, 1992).

3.5.4. Effect on cost efficiency

There is very limited data on cost comparisons between CAPI and paper-and-pencil personal interviews. Bond (1991) states that even when computers are used frequently in the fieldwork it will take about a year before the investment starts to pay back. Besides frequency of use, sample size is also a key factor for cost efficiency. Only with large sample sizes are the cost savings in printing, despatch, and data entry and editing (back-end costs) greater than the extra costs of questionnaire design and implementation (front-end costs). For example a long interview with closed questions only using a sample of 2000 or more will lead to a savings of 30%, a shorter questionnaire with a couple of open-ended questions and a sample of around 200 will only save around 5% (Bond, 1991). In these cost calculations the initial investment in equipment and in special training of staff has been excluded.

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 (total sample size CAPI+PAPI: 200). Their conclusion is that overall CAPI was more expensive, mostly because of added costs in training and supervising interviewers. In a larger experiment (around 300 respondents in each condition) 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 (1990) extrapolates these findings and concludes that when 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. However, several key-cost elements will decline as organisations gain experience in computer-assisted interviewing and hardware costs continue to fall.

3.6. Computer-Assisted Telephone Interviewing (CATI)

3.6.1. Effect on the respondent

In telephone interviewing the respondent will generally not notice whether a computer is used or not, therefore we may expect little, if any, difference between traditional telephone interviewing and CATI. This is confirmed by comparative studies. Groves and Nicholls (1986) conclude in a review that there are no differences in non-response, a conclusion also reached in comparative studies by Catlin and Ingram (1988) and Groves and Mathiowetz (1984).

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 (Catlin & Ingram, 1988; Groves & Nicholls, 1986; Weeks, 1992). This is similar to results found in the comparisons of traditional versus computer-assisted face-to-face interviewing reviewed above.

3.6.2. Effect on the interviewer

There is little research on the effect of CATI on the interviewers. Groves and Nicholls (1986) report that interviewers generally have a positive attitude toward CATI. They remark that acceptance of CATI strongly depends on the speed and reliability of the CATI system that is employed.[6] Weeks (1992) concludes that modern CATI systems are fast and reliable, and that interviewers prefer CATI to paper-and-pencil methods. Spaeth (1987) in her survey of survey organisations also reports that staff members in general (both supervisors and interviewers) preferred CATI above PATI.

Computer-assisted interviewing often leads to a greater standardisation of the interview, to the extent that interviewers sometimes complain about 'rigidity' (Riede & Dorn, 1991, p. 51). In general researchers will appreciate this greater standardisation, because it minimises interviewer bias (Fowler, 1991). Furthermore both Spaeth (1987) and Berry and O'Rourke (1988) reported that survey organisations tend to spend more time training interviewers for CATI than for PATI, and sometimes also employ more supervisory staff. There is some confirmation of greater standardisation 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 method.

[6] At the time of their review, CATI systems were running on mainframe or minicomputer systems, and both system speed and reliability were not always optimal for interactive interviewing. Modern microcomputer networks are much better in this respect.

3.6.3. Effect on data quality

Although CATI is the first form of computer-assisted interviewing that came into general use, there is little research on the influence of CATI on the data quality. In their review Groves and Nicholls (1986) conclude that CATI leads to less missing data because it prevents routing errors, but this effect is only important with complex questionnaires. For the same reason, post hoc data cleaning finds more errors with traditional paper-and-pencil methods than with CATI. They find no difference in respondent induced missing data because of 'don't know' and 'no answer' responses.

More recent research by Catlin and Ingram (1988) confirms these conclusions. Catlin and Ingram paid special attention to the possible effects on open questions; they found no differences in typing errors, codability or length of answer (number of words used). This is similar to results found in CAPI (cf. Baker, 1992).

3.6.4. Effect on cost efficiency

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

3.7. Computer-Assisted Self Interviewing (CASI)

Computer-assisted self-administered questionnaires are a relatively new development. CASI differs clearly from both CAPI and CATI by employing a different interviewing situation. The computer has taken the role of the interviewer. Theoretically, this combines the advantages of traditional self-administered questionnaires, such as more openness with sensitive questions, with the possibility of using complex question structures.

A disadvantage of CASI is that at present only selected populations can be studied. Comparative research on CASI has also mostly been done on selected populations, which either had access to computers, or received a computer for the duration of the study.

3.7.1. Effect on the respondent

Respondents generally like CASI; they find it interesting, easy to use, and amusing (Zandan & Frost, 1989; Witt & Bernstein, 1992). Beckenbach (1992) reports that more than 80% of the respondents

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

The general positive appreciation of CASI also shows in the relative high response ratio with Disk-by-Mail (DBM) surveys, and in the low panel mortality in the tele-interview. Saris (1989) reports for a large Dutch panel a mean response of the active panel members of 98% per week and a panel mortality of 15% per year. (However, the initial non-response for the panel was 50%; cf. Kalfs, 1993.) DBM response ratio's vary between 25% and 70%, and it is not unusual to have response ratio's of 40 to 50 percent without using any reminders (Salzman, 1992). 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 (Dillman, 1978; Heberlein & Baumgartner, 1978). Of course, one should realise that DBM is restricted to special populations who have access to a personal computer.

3.7.2. Effect on data quality

Respondents are generally positive about CASI. We expect that respondents will experience a higher degree of privacy and anonymity, which should lead to less social desirability bias. Beckenbach (1992) reports the two following controlled laboratory experiments that compare traditional self-administered questionnaires with CASI. In the first, Martin and Nagao (1989) compared CASI with face-to-face interviewing and a traditional self-administered questionnaire; using the Crowne-Marlowe social desirability scale they found less social desirability bias in the self-administered questionnaire, and even less with CASI. In the second, Evan and Miller (1969) compared CASI with a traditional self-administered questionnaire; they found that CASI leads to more openness with questions that are perceived as threatening, and no difference on non-threatening questions.

The same picture emerges in studies of electronic mail questionnaires. Sproull and Kiessler (1991) report about five experiments on decision making in small groups. Using an electronic network for communication leads to more open communication, more ideas, and general participation in the discussion. In the face-to-face situation the discussion tended to be dominated by one or two high-status individuals. This may also be the result of differences in the social interaction. However, in a direct comparison of a mail questionnaire and an electronic mail health-questionnaire Kiessler and Sproull (1986) also found fewer socially desirable answers in the electronic version. They also investigated other aspects of data quality in this study. Both the item non-response and the number of errors were lower with CASI. The responses to open questions did not differ, until the edit facilities of the CASI program were improved; then CASI led to longer and more personal answers.

The effect of computerisation on the quality of the data in self-administered questionnaires has also been a concern in psychological testing. The American Psychological Association's *Guidelines for computer-based tests and interpretations* (APA, 1986, p. 18) explicitly states that "… the

equivalence of scores from computerized versions should be established and documented before using norms or cutting scores obtained from conventional tests." The growing popularity of computerised psychological testing has led to several studies that assess the equivalence of conventional psychological tests and their computerised versions. In general, no differences between computer-assisted and paper-and-pencil tests were found in reliability and validity of the tests (Harrel & Lombardo, 1984; Parks et al., 1985). One study (Canoune & Leyhe, 1985) found that questions involving social pressure (conformity, evaluation) were answered differently in computerised and face-to-face questioning, with the face-to-face version leading to more social desirable answers and more tension reported by respondents, but other studies (Koson et al., 1970; Rezmovic, 1977) did not find this effect. A meta-analysis of 29 studies comparing conventional and computerised cognitive tests (Mead & Drasgow, 1993) found that power tests (ability tests without restrictive time limits) were highly equivalent (the cross-mode correlation is .97), but speed tests (cognitive tests measuring cognitive processing speed) were less equivalent (the cross-mode correlation is .72). Mead and Drasgow interpret the mode-effect for speeded tests as an effect of the importance of perceptual and motor skills in responding quickly to such tests.

The general conclusion is that paper-and-pencil and computer-assisted psychological tests are highly equivalent. This conclusion is corroborated by a study by Helgeson and Ursic (1989), who conclude from protocol analyses that there are no clear differences in the cognitive processes employed in responding to a traditional or a computer-assisted psychological test. The differences between the computerised and paper-and-pencil tests are the result of differences in motor skills under time pressure and possibly in less inhibition of respondents on highly sensitive topics.

3.7.3. Effect on cost efficiency

There are no systematic cost comparisons for CASI. The literature about Disk-by-Mail reports that DBM is generally more expensive than a comparable paper-and-pencil mail survey. However, the gain in response in a single mailing is thought to be worth the extra costs (e.g. Wilson, 1989).

3.8. Discussion

In reviewing the literature about CADAC, a conspicuous feature is that almost all publications start with summing up the potential advantages, such as cost reduction, higher efficiency, and improvement of response rate (cf. Snijkers, 1992). However, for most of these potential advantages the empirical evidence is still limited.

Most of the research on mode effects investigates the acceptability by respondents and some aspects of data quality. A systematic comparison of costs is difficult (see Groves, 1989), and consequently these are rare. When the total costs of paper-and-pencil and computer-assisted survey

research are compared, the evidence for cost reduction is not very strong (Baker, 1990; Catlin & Ingram, 1988; Nicholls & Groves, 1986). However, once the initial investments in equipment and personal are made, the technical possibilities and the efficiency of computer-assisted techniques of CAI will be preferred by both executives and clients (cf. Samuels, 1994). Just as CATI is now routine in telephone surveys, CAPI will become the future standard in market research.

There is little evidence that CATI or CAPI improves the response rate. Conversely, there is also no evidence for a decrease in response rate. In panel research (CAPAR) and in Disk-by-Mail surveys (DBM) there are advantages in the form of less panel attrition in CAPAR and better response rates in DBM. However, all relevant studies used selected groups of respondents.

At present, there is still little empirical research into the effect of computerised interviewing on data quality. Instead, most studies have investigated the acceptance of the new techniques by interviewers and respondents. Both interviewers and respondents evaluate computer-assisted interviewing positively. Comparative research has also paid little attention to the effect of computerisation on interviewer variance and on other aspects of interviewer behaviour (an exception is Groves & Mathiowetz, 1984).

Computerised methods of data collection generally have a positive effect on data quality. The improvements in data quality are similar for CAPI and CATI, and for CASI the improvements reported point in the same direction. Overall, there are advantages in using the computer with sensitive questions; respondents are less inhibited, but the differences are small. Much of the improvement of the data quality is the result of fewer interviewer and respondent errors; a well-designed computerised questionnaire simply does not allow certain types of mistakes. For example, one general finding is that item non-response decreases because there are no more routing errors.

A strong feature of CADAC is the potential to prevent errors by controlling routing and executing range and consistency checks, but CADAC is not being used to its full potential. The various aspects of data quality that have been studied are too limited. The strength of computer-assisted data collection methods is the ability to increase the power of interviewing and thus be able to answer more complex research questions (Bond, 1991). We should explore the potential of the computer and use techniques for data collection that are impossible or impractical without it. An example is the application of computerised diaries in time-budget research, which makes it possible to use very fine distinctions (Kalfs, 1993). Another possibility is the presentation of graphical stimuli, for instance for magnitude scaling (Saris, 1989), and automatic time registration in interviews (Bassili & Fletcher, 1991). Bond (1991) gives a detailed description of two measurement techniques that would be almost impossible to use without a computer: Natural Grouping and Adaptive Conjoint Analysis. These methods are extremely useful for brand research and positioning of new products.

Another interesting capability of computerised interviewing is extreme tailoring. In this technique different respondents are asked different subsets of the questionnaire. There is a complex routing

structure, in which the answers to (combinations of) earlier questions determine the questions that are asked next. This allows for a high degree of structured individualisation of the questionnaire without the subjectivity that comes with an open interview. Kuijlen (1993) has used extreme tailoring in eliciting scenarios that respondents use in purchasing mortgages. A related method is adaptive testing used in psychological measurement. In computerised adaptive testing the computer program selects questions from a pool of possible questions related to the psychological trait to be measured. Based on the responses to questions asked earlier, the program calculates a provisional estimate of the respondent's score, and then selects from the pool of remaining questions that question that maximises the information than can be obtained from the next question. The process ends when all questions have been used, or when the program decides that the respondent's score can be estimated with sufficient precision.

Instead of tailoring the questions to the respondents, respondents can also be tailored to the questions. Poulsen (1993) describes an adaptive CATI system that segments respondents during the interview with a minimum of questions. The resulting segmentation variable is then used in the remainder of the interview for skipping and branching. Thus, different subgroups of respondents receive different questions.

Clearly CADAC is no panacea for good data quality. Computerised data collection methods require one to do almost everything that is needed with a good paper-and-pencil interview, and to add extra efforts in computer implementation, in testing the questionnaire, and in extra interviewer training. This investment is earned back in far less interviewer error and the error-free administration of complex questionnaires. We should avoid strong time pressure on both interviewers and respondents: time pressure interacts with the perceptual and motor skills needed for correct reading of questions and typing in the answer. In addition, research has shown that question texts are harder to read on a monitor than on paper (Mead & Drasgow, 1993), which implies that ergonomical text presentation and careful screen design is important for good CADAC surveys. Only if such efforts are made, is there the opportunity for obtaining better data. But, to paraphrase Bond (1991) we have to push further and increase the value of computer interviewing. Computer-assisted data collection has given us beautiful tools to work with. We should use these tools intelligently and devise new ways to measure what respondents can tell us if we only 'listen'.

References

APA (American Psychological Association), 1986, Guidelines for computer-based tests and interpretations. (American Psychological Association, Washington, DC.)

Argyle, M., and J. Dean, 1965, Eye-contact, distance and affiliation. (Sociometry, Vol. 28, pp. 289-304.)

Baker, R.P., 1990, What we know about CAPI: Its advantages and disadvantages. (Paper presented at the annual meeting of the American Association of Public Opinion Research, Lancaster, Pennsylvania.)

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

Baker, R.P., and N.M. Bradburn, 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).

Bassili, J.N., and J.F. Fletcher, 1991, Response time measurement in surveys. (Public Opinion Quarterly, Vol. 55, pp. 331-346.)

Beckenbach, A., 1992, 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. (In German: Befragung mit dem Computer, Methode der Zukunft?. Anwendungsmöglichkeiten, Perspektiven und experimentelle Untersuchungen zum Einsatz des Computers bei Selbstbefragung und persönlich-mündlichen Interviews. Ph.D. thesis. University of Mannheim, Mannheim, Germany.)

Bemelmans-Spork, M., F. Kerssemakers, D. Sikkel, and H. Van Sintmaartensdijk, 1985, Report of an experiment on the use of portable computers in person and household surveys. (In Dutch: Verslag van het experiment 'het gebruik van draagbare computers bij persoons en gezinsenquêtes'. Statistics Netherlands, Voorburg.)

Bennet, D., and C. Goodger, 1993, Interviewer training for CAI at OPCS. (Paper presented at the 1993 conference of the Study Group on Computers in Survey Analysis. City University, London.)

Berry, S.H., and D. O'Rourke, 1988, Administrative designs for centralized telephone survey centers: implications of the transition to CATI. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 457-474.)

Biemer P.P., R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds.), 1991, Measurement Errors in Surveys. (Wiley, New York.)

Blyth, W., and H. Piper, 1994, Speech recognition - a new dimension in survey research. (Journal of the Market Research Society.)

Bond, J., 1991, Increasing the value of computer interviewing. (In: Proceedings of the 1991 ESOMAR congress: Trends in data collection and analysis.)

Borg, I., and P. Mohler (eds.), 1994, Trends and perspectives in empirical social research. (Walter de Gruyter, Berlin.)

Bradburn, N.M., M.R. Frankel, R.P. Baker, and M.R. Pergamit, 1992, A comparison of CAPI with PAPI in the NLS/Y. (Information Technology in Survey Research Discussion Paper 9. NORC. Chicago.) (Also presented at the 1991 AAPOR conference, Phoenix, Arizona.)

Canoune, H.L., and E.W. Leyhe, 1985, Human versus computer interviewing. (Journal of Personality Assessment, Vol. 49, pp. 103-106.)

Catlin, G., and S. Ingram, 1988, The effects of CATI on costs and data quality: A comparison of CATI and paper methods in centralized interviewing. (In: Groves et al. (eds.), Telephone Survey Methodology, pp. 437-456.)

Couper, M.P., and R.M. Groves, 1992, Interviewer reactions to alternative hardware for computer assisted personal interviewing. (Journal of Official Statistics, Vol. 8, pp. 201-210.)

De Leeuw, E.D., 1989, Computer assisted data collection; A review of technological changes. (In Dutch: Computergeleid enquêteren; Een overzicht van technologische vernieuwingen. Tijdschrift voor Onderwijs Research, Vol. 4, pp. 201-213.)

De Leeuw, E.D., 1993, Data quality in mail, telephone and face to face surveys. (Ph.D. Thesis. TT-Publikaties. Amsterdam.)

Denny, M., and L. Galvin, 1993, Improved quality at the touch of a button; The use of computers for data collection. (In: Proceedings of the 1993 Market Research Society Conference.)

Dillman, D.A., 1978, Mail and Telephone Surveys: The Total Design Method. (Wiley, New York.)

Dillman, D.A., 2000, Mail and Internet Surveys: The Tailored Design Method. (Wiley, New York.)

Edwards, B., D. Bittner, W.S. Edwards, and S. Sperry, 1993, CAPI effects on interviewers: A report from two major surveys. (Paper presented at the U.S. Bureau of the Census Annual Research Conference, Washington D.C.)

Evan, W.M., and J.R. Miller, 1969, Differential effect of response bias of computer vs. traditional administration of a social science questionnaire: An exploratory methodological experiment. (Behavioral Science, Vol. 14, pp. 216-227.)

Faulbaum, F. (ed.), 1992, Advances in Statistical Software 3. Softstat 1991. (Fischer, Stuttgart.)

Fowler, F.J., 1991, Reducing interviewer-related error through interviewer training, supervision, and other means. (In: Biemer et al. (eds.), Measurement Errors in Surveys, pp. 259-278.)

Gladitz, J., and K.G. Troitzsch (eds.), 1990, Computer aided sociological research. (Akademie-Verlag, Berlin.)

Groves, R.M., 1989, Survey Errors and Survey Costs. (Wiley, New York.)

Groves, R.M., P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (eds.), 1988, Telephone Survey methodology. (Wiley, New York.)

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

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

Havis, M.J., and M.J. Banks, 1991, Live and automated telephone surveys: A comparison of human interviewers and an automated technique. (Journal of the Market Research Society, Vol. 33, pp. 91-102.)

Harrel, T.H., and T.A. Lombardo, 1984, Validation of an automated 16PF administration procedure. (Journal of Personality Assessment, Vol. 48, pp. 216-227.)

Heberlein, T.A., and R. Baumgartner, 1978, Factors affecting response rates to mailed questionnaires; A quantitative analysis of the published literature. (American Sociological Review, Vol. 43, pp. 447-462.)

Helgeson, J.G., and M.L. Ursic, 1989, The decision process equivalency of electronic versus pencil-and-paper data collection methods. (Social Science Computer Review, Vol. 7, pp. 296-310.)

Higgins, C.A., T.P. Dimnik, and H.P. Greenwood, 1987, The DISKQ survey method. (Journal of the Market Research Society, Vol. 29, pp. 437-445.)

Hippler, H.J., and A. Beckenbach, 1992, Computer-assisted questioning: Developments, restrictions and future possibilities. (In: Faulbaum (ed.), Advances in Statistical Software 3, pp. 403-410.)

Hox, J.J., S. De Bie, and E.D. De Leeuw, 1990, Computer-assisted (telephone) interviewing: a review. (In: Gladitz et al. (eds.), Computer aided sociological research.)

Jacobs, M.A., 1993, Buying or copying software? A study of pc-users. (In Dutch: Software kopen of kopiëren? Een sociaal-wetenschappelijk onderzoek naar pc-gebruikers. Thesis-publishers Amsterdam.)

Kalfs, N., 1993, Hour by hour; Effects of the data collection mode in time use research. (Nimmo, Amsterdam.)

Kasprzyk, D., G.J. Duncan, and G. Kalton, 1989, Panel surveys. (Wiley, New York.)

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

Koson, D., C. Kitchen, M. Kochen, and D. Stodolosky, 1970, Psychological testing by computer. The effect of response bias. (Educational and Psychological Measurement, Vol. 30, pp. 803-810.)

Kuijlen, A.A.A., 1993, The scenario method: A study of complex consumer decisions using computer assisted interviewing. (In Dutch: De scenario benadering. Een onderzoek naar complexe consumentenbeslissingen met behulp van computergestuurd enquêteren. Ph.D. thesis. Katholieke Universiteit Brabant, Tilburg, The Netherlands.)

Martin, C.L., and D.H. Nagao, 1989, Some effects of computerized interviewing on job applicant responses. (Journal of Applied Psychology, Vol. 74, pp. 72-80.)

Martin, J., C. O'Muircheartaigh, and J. Curtice, 1993, The use of CAPI for attitude surveys: An experimental comparison with traditional methods. (Journal of Official Statistics.)

Mead, A.D., and F. Drasgow, 1993, Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. (Psychological Bulletin, Vol. 114, pp. 449-458.)

Nicholls, W.L. II, and R.M. Groves, 1986, The status of Computer-Assisted Telephone Interviewing: Part 1- Introduction and impact on cost and timeliness of survey data. (Journal of Official Statistics, Vol. 2, pp. 93-115.)

Olsen, R.J., 1992, The effects of computer assisted interviewing on data quality. (Paper presented at the fourth Social Science Methodology Conference, Trento, Italy.)

Parks, B.T., D.E. Mead, and B.L. Johnson, 1985, Validation of a computer administered marital adjustment test. (Journal of Marital and Family Therapy, Vol. 11, pp. 207-210.)

Porst, R., M. Schneid, and J.W. Van Brouwershaven, 1994, Computer-assisted interviewing in social and market research. (In: Borg et al. (eds.), Trends and perspectives in empirical social research.)

Poulsen, C.S., 1993, Artificial intelligence applied to computer aided interviewing. (In: Proceedings of the ESOMAR symposion on information based decision making in marketing, 1993, Paris.)

Rezmovic, V., 1977, The effects of computerized experimentation on response variance. (Behaviour Research Methods & Instrumentation, Vol. 9, pp. 144-147.)

Riede, T., and V. Dorn, 1991, Acceptance of laptops for household surveys in Germany. (In German: Zur Einsetzbarkeit von Laptops in haushaltsbefragungen in der Bundesrepublik Deutschland. Heft 20 der Schriftenreihe Ausgewählte Arbeitsunterlagen zur Bundesstatistik. Statistisches Bundesamt, Wiesbaden, Germany.)

Roos, M., L. Jaspers, and G. Snijkers, 1999, The Survey on Short Term Economical Indicators via the Internet. Data collection via E-mail: A first Study. (In Dutch: De Conjunctuurtest via Internet. Gegevensverzameling met e-mail: een eerste proef. CBS report, BPA no. H4350-99-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Roos, M., and G. Snijkers, 1999, Blaise IS under Construction. Data Collection via the Internet: A second study. (In Dutch: Blaise IS in de Bouw. Gegevensverzameling via het Internet: de tweede proef. CBS report, BPA no. H4435-99-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Saltzman, A., 1992, Improving response rates in disk-by-mail surveys. (In: Sawtooth software conference proceedings. Sawtooth software, Evanston.)

Samuels, J., 1994, From CAPI to HAPPI: A scenario for the future and its implications for research. (In: Proceedings of the 1994 ESOMAR congress: Applications of new technologies.)

Saris, W.E., 1989, A technological revolution in data collection. (Quality & Quantity, Vol. 23, pp. 333-349.)

Saris, W.E., 1991, Computer-assisted interviewing. (Quantitative Applications in the Social Sciences Series, No. 80. Sage, Newbury Park.)

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

Snijkers, G. 1992, Computer-Assisted Interviewing: Telephone or Personal ? A Literature Study. (In: Westlake et al. (eds.), Survey and Statistical Computing, pp. 137-146.)

Spaeth, M.A., 1987, CATI facilities at survey organizations. (Survey Research, Vol. 18, 18-22.)

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

Sproull, L., and S. Kiessler, 1991, Computers, networks and work. (Scientific American, pp. 84-91.)

Statistics Sweden, 1989, Computer assisted data collection in the labour force surveys: Report of technical tests. (Statistics Sweden, Stockholm.)

Thornberry, O., B. Rowe, and R. Biggar, 1991, Use of CAPI with the U.S. National Health Interview Survey. (Bulletin de Méthodologie Sociologique, No. 30, pp. 27-43.)

Van Bastelaer, A.M.L., F. Kerssemakers, and D. Sikkel, 1987, A test of the Netherlands continuous labour force survey with hand-held computers: Interviewer behaviour and data quality. (In: Automation in survey processing. CBS Select 4. Staatsuitgeverij, Den Haag.)

Van Doorn, L., 1987, The use of microcomputers in panelsurveys. (In Dutch: Het gebruik van microcomputers in panelonderzoek. Jaarboek van de Nederlandse Vereniging van Marktonderzoekers 1987-88 (Yearbook of the Dutch Association for Market Research), pp. 9-21.)

Van Doorn, L.T., and T. Hess, 1987, Keeping panel respondents fit and willing: Using computerized interviewing in panel research to decrease the burden of respondents and improve the quality of the data. (In: Proceedings of the 1989 ESOMAR congress.)

Waterton, J.J., 1984, Reporting alcohol consumption: The problem of response validity. (In: Proceedings of the section on survey research methods of the American Statistical Association, pp. 664-669. (ASA, Washington D.C.)

Waterton, J.J., and J.C. Duffy, 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, Vol. 2, pp. 173-182.)

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

Westlake, A., R. Banks, C. Payne, and T. Orchard (eds.), 1992, Survey and Statistical Computing. (North-Holland, Elsevier Science Publishers, Amsterdam.)

Wilson, B., 1989, Disk-by-Mail Surveys: Three Years Experience. (In: Sawtooth software conference proceedings. Sawtooth software, Evanston.)

Wings, H., and G. Snijkers, 1998, Exploring Data Collection by the Internet. (CBS report, BPA no. H5117-98-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.) (Paper presented at the 5th International Blaise Users Conference, November 11-13, 1998, Lillehammer, Norway.)

Witt, K.J., and S. Bernstein, 1992, Best practices in disk-by-mail surveys. (In: Sawtooth software conference proceedings. Sawtooth software, Evanston.)

Woijcik, M.S., S. Bard, and E. Hunt, 1992, Training field interviewers to use computers: a successful CAPI training program. (Information Technology in Survey Research Discussion Paper 8. NORC, Chicago.) (Also presented at the 1991 AAPOR conference, Phoenix, Arizona).

Zandan, P., and L. Frost, 1989, Customer satisfaction research using disk-by-mail. (In: Sawtooth software conference proceedings. Sawtooth software, Evanston.)

Chapter 4
Methods 1

Cognitive Laboratory Methods: Current Best Practices

Summary: *At the Questionnaire Laboratory of Statistics Netherlands several methods for pre-testing questionnaires are being used, to investigate the question-and-answer process. These methods are: expert (re)appraisal, focus groups, in-depth interviews, and behavioural coding. Within in-depth interviews techniques like thinking aloud, follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes can be used. In this chapter these methods and techniques will be discussed from a practical point of view, i.e. how they are applied in practice. This chapter provides an overview of current best practices. (In the chapters to come, case studies in which these methods have been applied, will be discussed.)*
These methods are presented within the context of a 5-step (pre-)test model for survey data collection development. In 5 steps the data collection procedures are developed and pre-tested, shifting from testing a first draft of the questionnaire to testing the data collection procedures in the field, and implementation of the survey. For each step, the model indicates what methods can be used. In a full pre-test program, following this model and the guidelines (including scientific standards) for application presented in this chapter, all aspects of the survey will be carefully tested in advance. This model offers a methodology to shift research findings from the laboratory to the field.

Keywords: *Methods for Pre-testing Questionnaires, Question-and-Answer Process, Expert (Re)appraisal, Focus Group, In-depth Interview, Behavioural Coding, 5-Step (Pre-)test Model for Data Collection Development.*

4.1. Introduction: A (pre-)test model for survey data collection development

When designing a survey, many aspects have to be taken care of like the sample, the interviewing mode, interviewing procedures, data cleaning and weighting procedures. Another one is the questionnaire, which is ideally pre-tested in a cognitive laboratory. However, cognitive laboratory methods may not only be used to pre-test questionnaires; they may also be used for evaluating and monitoring data collection procedures, associated with other sources of measurement errors like mode and interviewer aspects, in order to reduce or prevent these errors from occurring.

At the Questionnaire Laboratory at Statistics Netherlands a coherent model for designing and (pre-)testing data collection procedures (including questionnaires) has been developed. In this 5-step model, as described by Akkerboom and Luiten (1996), Akkerboom and Dehue (1997), Dehue (1997), and Akkerboom, Dehue and Snijkers (1998), a relation is presented between the design stage and the (pre-)testing methods that can be used. With every step, the focus gradually shifts from prototypes of the questionnaire and the data collection procedure to implementation of the survey in the field. The methods to be used in each step are aimed at improving the quality of the questionnaire and the efficiency of the data collection procedures, by (1) investigating the errors that might occur, (2) getting insight in why they occur, and (3) designing adaptations to mend those errors in the questionnaire and the data collection procedures. The model is presented in table 4.1. In table 4.2 a short definition of each method is listed.

The first step in the development process of a survey is a definition and feasibility study. In this step the questionnaire and the data collection procedure are developed and reviewed in order to see whether the survey is feasible. Here, questions are asked like: ‘What data have to be gathered and how should this be done?’, and ‘Are these data measurable by interviewing?’ When designing a questionnaire, the designer (or group of designers) starts with the research objectives and concepts to be measured, followed by operationalisations (Schwarz, 1997). Hox (1997) discusses two research strategies to operationalise concepts: the theory driven or ‘top down’ strategy, which starts with constructs and works towards observable variables, and the data driven or ‘bottom up’ strategy, which starts with observations and works towards theoretical constructs. As for the top down approach there is no questionnaire present at the start, while for the bottom up approach we start with a questionnaire already present and find out whether the concepts to be measured are operationalised appropriately. The results of this step are prototypes of a questionnaire and of a data collection procedure, or in case the survey does not seem to be feasible, a ‘no-go’ decision.

In the second step the questionnaire is pre-tested in a cognitive laboratory in order to investigate the question-and-answer process: ‘Is the questionnaire a valid measuring instrument?’ In this step, typical cognitive laboratory methods like in-depth interviews and focus groups are used. These methods will be discussed in this chapter, from a practical point of view, i.e. how they are applied. (Apart from investigating the question-and-answer process, in this phase also other aspects of the questionnaire are tested, like the readability of questions by interviewers,

Table 4.1. A 5-step (pre-)test model of data collection development (*)

Step	Result of step	Topics to be addressed	(Pre-)test methods	
			to be used	(pre-)test size*
1. Definition/ feasibility study	Prototypes of: • questionnaire, • data collection procedure • data processing procedure	Designing: • questionnaire, • data collection and • data processing procedure: What data have to be gathered how?	• Review of literature, research papers, meta-analysis • Expert appraisal • Exploratory focus groups	 • *e* = 3 – 10 • *u*, *e*, *i*, *v* = (1x5) – (3x10)
2. Qualitative laboratory test	'Less error-prone' questionnaire, by revision of: wording, sequence of questions, inter-viewer procedures, interviewing mode	Pre-testing ques-tionnaire (question-and-answer process; readability of questions; usability of questionnaire and duration of interview)	• Expert (re)appraisal • Focus groups • In-depth interviews • Observations (monitoring standardised interviews)	• *e* = 3 – 10 • *v*, *i*, *u*, *e* = (1x5) – (3x10) • *v* = 5 – 50 • *v*, *n* = 5 – 50 *i* = 1 – 3
3. Qualitative operational field test	'Less error-prone': • questionnaire, • data collection and • data processing procedure	Pre-testing: • questionnaire (measurement quality), and • data collection and • data processing (process efficiency) in the field	• Monitoring standardised interviews: evaluation/test questions, observations • Focus groups / debriefings • Re-interviews • Expert (re)appraisal • (Item) non-response analysis	• *n* = 25 – 300 *i* = 5 – 20 • *n*, *i* = (1x5) – (3x10) • *n* = 25 – 100 *i* = 5 – 20 • *e* = 3 – 10
4. Quantitative pilot study	Final questionnaire, data collection and data processing procedure	Testing data collection and data processing procedure in the field: costs and benefits	• Monitoring standardised interviews: evaluation/test questions, observations • (Item) non-response analysis • Data analysis (external validity) • Controlled statistical experiments • Other monitoring methods: re-interviews, focus groups/ debriefings	• *n* =200–500 • *n* = 100–250 *n*, *i* = (1x5) – (3x10)
5. Implemen-tation	Get survey going: all preparations for carrying out the survey have been made	Implementation of final questionnaire, data collection and data processing procedure	• Monitoring standardised interviews in survey: evaluation/test questions, observations • Other monitoring methods: re-interviews, focus groups/debriefings • (Item) non-response analysis, data analysis	• *n* = sample size • *n* = 100–250 *n*, *i* = (1x5) – (3x10)

* The pre-test size indicates the type and number of participants involved: *v* = volunteer respondent, *n* = sample respondent, *e* = cognitive laboratory expert, *i* = interviewer, *u* = data user (client, researcher).
(*) Adapted from: Akkerboom & Dehue (1997: table 1a, p. 129).

the usability of the questionnaire and the duration of an interview. See chapter 10.) Ideally, in this phase the questionnaire is improved in an iterative process of pre-testing and adapting.

Table 4.2. Overview of methods (*)

(Pre-)testing method	Definition
Review (step 1)	Risk assessment based on (meta-)data from the past. This is not a proper pre-testing method, but a review of earlier evidence for measurement error risks of question and data collection procedures. For example: previous (pre-)tests, known response rates for various data collection modes, statistical analysis of previous surveys on item and unit non-response rates, subject matter-knowledge.
Expert (re)appraisal (steps 1 - 3; subsection 4.2.1)	Risk assessment based on judgement by experts (with background in fieldwork design, questionnaire design or cognitive methods). This is not a proper pre-testing tool since no respondents are involved, but an exploratory judgement by experts on questionnaire and data collection procedure. Expert judgement may provide the proper focus for pre-tests to follow.
Focus group / debriefing (steps 1 - 5; subsection 4.2.2)	An open group interview with a topic list as agenda, resulting in an overview of general information. Participants are experts, survey data users, interviewers, or respondents. Step 1 focus groups are mainly exploratory; step 2 to 5 focus groups are evaluative (debriefings).
In-depth interviews (step 2; subsection 4.2.3)	An open interview in which a (self- or interviewer-administered) questionnaire is tested in a laboratory setting with one respondent and an interviewer, in which several techniques are used to provide an in-depth investigation of the question-and-answer process. Techniques that may be applied are: thinking aloud, follow-up probing, meaning oriented probing, paraphrasing, targeted test/evaluation questions, and vignettes. The interview is guided by risk hypotheses on the question-and-answer process, coming from reviews, expert (re)appraisal, focus groups, and observations.
Observations (steps 2 - 5; subsection 4.2.4)	Observation of standardised survey interview by interviewer or experts. Observation is an exploratory way to monitor standardised interviews e.g. by means of predetermined codes associated with respondent and interviewer behaviour.
Re-interview (steps 3 - 5)	Re-contacting a sub-sample of the (non-)respondents with an additional standardised interview, after the field interviews have been conducted. Re-interviews are guided by risk hypotheses to research or monitor interviewer or respondent characteristics or behaviour in the field (e.g. unit and item non-response, understanding of situations using vignettes).
(Item) non-response analysis (steps 3 - 5)	Exploratory analysis of unit and item non-response to monitor interviewer performance (to identify dropouts), and data quality with regard to respondent and questionnaire characteristics (to identify badly performing sub-groups of respondents e.g. regional dropouts, and badly performing questions).
Data analysis (external validity) (steps 4 - 5)	Statistical comparison of target variables with other data sources, in order to research correspondence (co-variation) of results. (Cook and Campbell (1979, p. 39) refer to external validity as “the approximate validity with which conclusions are drawn about the generalizability of a causal relationship to and across alternate measures, populations of persons, settings, and times.”)
Experiments (step 4)	Controlled statistical comparison of design variants (in questionnaire and/or procedure), guided by risk hypotheses. Examples: split-ballot experiment, test re-test experiment, multi-trait multi-method approach. (See e.g. Cook & Campbell, 1979; Saris, 1998; Van den Brakel, 2001.)

(*) Adapted from: Akkerboom & Dehue (1997: table 1b, pp. 130-131).

The third step is a qualitative operational test in which the improved questionnaire and the data collection procedure are tested in the field, in order to find out what major problems may be faced in the field: ‘Do the questionnaire and the data collection procedure work out well, or does the gathering of data have to organised in another way?’ Actually, in this step the cognitive laboratory is ‘moved into the field’. In the fourth step a large-scale quantitative pilot study is carried out to investigate costs and benefits of the survey, resulting in the final questionnaire and data collection procedure. Going from step 2 to steps 3 and 4, the focus shifts from measurement quality to process efficiency, and accordingly, the methods listed shift from qualitative, exploratory ones to quantitative, confirmatory methods. The fifth step is the implementation of the survey. As a result

of this step all preparations have been made for getting the survey into the field. Although all procedures have been carefully tested when all steps in this model have been followed, in this step also some test methods are listed in order to continuously monitor the data and the procedures during execution of the survey.

The steps in the model follow a logical sequence in the development process of a survey, but in practice it may be an iterative process, in which (as a results of a next step) it is necessary to go backwards. For instance, as a result of step 3, the questionnaire has to be changed, making more laboratory tests necessary (step 2). In practice however, the development process is limited by money, people and time constraints. So, there may be an overlap between steps, e.g. with steps 2 and 3, testing both the questionnaire and the data collection procedure in the lab and in the field at the same time. Also, steps may be skipped. The steps that are in every development process are steps 1 and 5. However, the questionnaire is essential within the whole process, since it defines the data model. Therefore, testing a questionnaire in step 2 seems a necessary step before going on. As for steps 3 and 4, at a large data collection institution like Statistics Netherlands, data collection procedures are standardised. Furthermore, these steps take a lot of time and effort to carry out. Thus, steps 3 and 4 may be skipped, which however, makes an evaluation of the questionnaire and the data collection procedure in the field (in step 5) necessary. Thus, a minimal program for developing and testing may be formulated. In table 4.3 this minimal program is listed, including the necessary steps and the methods to be used. It may be clear however, that a better way to prepare a survey is to carefully pre-test all aspects of the survey in advance in stead of afterwards. This model offers a methodology to shift research findings from the laboratory to the field.

Table 4.3. Minimal program for developing and testing a data collection procedure

Step	(Pre-)test methods	
	to be used	(pre-)test size*
1. Definition/ feasibility study	• Review of literature, research papers • Expert appraisal	 • e = 3 – 5
2. Qualitative laboratory test	• Observations of standardised interviews • Focus groups	• v = 5 – 10; i = 1 • v = 1 x (5 – 10)
5. Implementation	• Monitoring methods: observations, re-interviews, focus groups/debriefings • (Item) non-response analysis, data analysis	• n = 100–250 n, i = 2 x (5 – 10)

* v = volunteer respondent, n = sample respondent, e = cognitive expert, i = interviewer.

In section 4.2 the pre-test methods, as mentioned in table 4.1, will be described. We will concentrate on methods as applied by the Questionnaire Laboratory at Statistics Netherlands to pre-test questionnaires: expert (re)appraisal, focus groups, in-depth interviews, and observations (or behavioural coding). Within in-depth interviews several targeted techniques have been applied: thinking aloud and follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes. Most of these methods and techniques are also used by other National Statistical Institutes, like the cognitive laboratories at the U.S. Census Bureau (USCB, 1998) and Statistics Canada (Gower, 1991; Gower et al., 1998).

As mentioned above, the pre-test methods in table 4.1 are grouped according to their use in the development process of a survey. Another taxonomy of pre-test methods is presented by Forsyth and Lessler (1991). They group the methods according to four general types of methods: expert evaluation (like behaviour coding and expert appraisal), expanded interviews (think-aloud interviews, follow-up probing among others), targeted methods (e.g. meaning-oriented probing, paraphrasing) and group methods (like focus group). However, we feel that this taxonomy may not be very applicable in practice, since it does not indicate when to use a method or what each method is aimed at. And furthermore, the groups of methods are not exclusive, e.g. targeted methods like meaning-oriented probing, may be used within expanded interviews like think-aloud interviews. Nevertheless, this taxonomy is useful in combination with the overview in table 4.1.

In yet another overview of methods, Sudman, Bradburn and Schwarz (1996) indicate what each method is aimed at. Their overview is based on the question-and-answer process as described by Tourangeau and Rasinski (1988): comprehension of wording and interpretation of response task, information retrieval, judgement or evaluation, and reporting. For each method they indicate what aspect of this process is investigated. Another classification based on the question-and-answer process is presented by Groves (1996). For some of these methods Campanelli (1997) gives a rating on usefulness (the number of problems uncovered), as based on several studies in which these methods have been compared (Campanelli, Rothgeb et al., 1991; Esposito et al., 1991; Fowler, 1992; Fowler & Roman, 1992; Presser & Blair, 1994).

In the next session, these aspects of the methods will also be indicated. However, in the literature these methods are described in a general way, while little is said about guidelines on how to apply these methods in practice. Therefore, section 4.2 concentrates on the latter, providing an overview of current best practices. Section 4.3 concludes this chapter with a summary of features of these methods, including scientific standards for application. In this summary all aspects of the overviews mentioned above are being combined.

4.2. Cognitive laboratory methods

4.2.1. Expert (re)appraisal

In an expert (re)appraisal the (draft) questionnaire is reviewed by a panel of cognitive experts (Lessler & Forsyth, 1996). Therefore, this method is also called ‘expert panel review’ (Presser & Blair, 1994) or ‘expert analysis’ (Forsyth & Lessler, 1991). It is meant to get expert opinions on the (draft) questionnaire in the design step (step 1 in table 4.1), the pre-testing step (step 2), or the operational test (step 3), in order to get an understanding of potential problems in the (draft) questionnaire with respect to the question-and-answer process, i.e. potential difficulties the respondent might have in understanding and answering the question. At the Questionnaire Laboratory at Statistics Netherlands, this method is very frequently used to give quick advice on

draft questionnaires, since the method is very time efficient. Expert (re)appraisal also serves as a means to design pre-testing protocols, i.e. identifying the issues that have to be tested in the laboratory (in step 2 with focus groups or in-depth interviews) or in the field (step 3).

It is common practice among researchers to have draft versions of a questionnaire reviewed by colleagues with a background in the topics addressed in the questionnaire, fieldwork design or questionnaire design. Within a cognitive laboratory, questionnaires may be reviewed from a cognitive perspective: the question-and-answer process. On the basis of this process, Forsyth, Lessler and Hubbard (1992) developed a coding scheme for expert appraisal. A current version of this 'Questionnaire Appraisal Coding System' is presented in table 4.4 (Lessler & Forsyth, 1996).

At the Questionnaire Laboratory at Statistics Netherlands, we found this coding scheme too detailed and extensive for effective and efficient use in assessing question errors. Furthermore, the information obtained is quantitative. The codes in table 4.4 only indicate that there might be a specific problem within a question (e.g. that there is a technical term used in the question); the codes do not indicate what the cause of the problem is (what term or word exactly is problematic) and how the question should be improved (in this example, how the question should be rephrased). Clients of the Questionnaire Laboratory did not only expect a classification of errors, but above all suggestions for improvement of the questions (with an explanation on why the original question might be troublesome). In order to get an advice that not only included possible errors, but also suggestions for revision, a condensed version of the system in table 4.4 was used for expert (re)appraisal (Snijkers et al., 1994).

This condensed scheme is presented in table 4.5. It lists only the most common errors. In addition to coding the expected problems, experts had to refer to the parts of the questions that are problematic according to them, as well as present suggestions for improvement. To get an inter-subjective appraisal, we had a questionnaire reviewed by at least three experts, going up to ten (when there was enough time; although, most of the time, questionnaire reviews had to be given within one week.). After all experts had appraised the questions individually, one of them collected the reviews and classified the observations. In case of inconsistencies the experts discussed the results, in order to get one conclusion. This method resulted in quick reviews, including both possible errors and suggestions for revision. The coding system in table 4.5 is also used to analyse in-depth interviews (see subsection 4.2.3, figure 4.1).

In 1997, Akkerboom and Dehue of the Questionnaire Laboratory, extended this system to the one in table 4.6. This system serves to classify errors in questions too. But, in contrast to the other systems, Akkerboom and Dehue did not order the errors according to the stages of the question-and-answer process. As Akkerboom and Dehue state (p. 142): "It is an eclectic scheme based on various sources in the literature (e.g. Foddy, 1993) and it is not meant to replace sophisticated coding schemes used in (cognitive) expert appraisal, compare Lessler & Forsyth (1996).

Table 4.4. Questionnaire Appraisal Coding System (*)

Comprehension		*Retrieval*	*Judgment*	*Response Generation*
Establish reference set boundaries. Establish reference period boundaries.		Retrieval of information.	Evaluate retrieved information.	Establish response set of boundaries. Generate response.
Reference set	*Reference set*			
Task requirements	*Potential problems cont'd*	*Task requirements*	*Task requirements*	*Task requirements*
Reference set type	*Questions cont'd*	*Retrieval task*	*Type of judgment process*	*Response description*
Current characteristic/behavior Past characteristic/behavior Current attitude/opinion Past attitude/opinion General knowledge	Question Structure: Hidden question Complex syntax Implicit assumption Several questions Several definitions Unclear goal Q/A mismatch Violates con coven	Remember episode Remember set of episodes Remember general information Remember previous answer Recall attitude	Estimate total Determine +/– occurrence Determine +/– match Determine date/onset Determine age Estimate duration Estimate average Complex estimation	Yes/no Qualitative: category Qualitative: ordinal Qualitative: open Quantitative: count Quantitative: complex Duration Time point Age Answer hidden question
Focus	*Problem summary*	*Memory process*	*Information integration*	*Potential problems*
Self-report Proxy report	Vague reference set Complex reference set Constant behavior/implement Nonsalient attitude	Recall Recognition Heuristic/inference Mixed above	Count Qualitative judgment Quantitative judgment	*Instructions* Hidden instructions Hidden definitions
Reference set level	*Reference period*	*Memory content*	*Potential problems*	*Responses*
Basic Subordinate Superordinate Multi-level	*Task requirements*	General self-knowledge General world knowledge Specific behavior (or try) Class of behavior Affect/attitude Time point/interval	*Information evaluation* Accuracy evaluation Sensitive behavior Sensitive attitude Sensitive (general) Socially undesirable	Terminology: Technical terms Undefined terms Ambiguous/vague terms Response structure: Boundary problem Overlapping categories Missing categories Non-dominant order
Reference set changes	*Specific period* Lifetime 12-months 30-days _____ Today Tied to behavior/prev. question Undefined: e.g., currently	*Potential problems* High detail Low detail Unexpected detail Shift: psychological reference period	*Consequence evaluation* Safety consequences Legal consequencs Social consequencs Behavioral consequences	
Domain change Level change Abrupt change Carry-over ref.set				
Potential problems			*Information/response congruence* Incongruent	
Instructions				
Conflicting instructions Inaccurate instructions Hidden instructions Complex syntax Unclear examples Unclear layout	*Reference period changes* Change from prior question: No change Wider Narrower Distal boundaries Proximal boundary			
Questions	*Potential problems*			
Technical Terms: Present Undefined Ambiguous or vague	Unanchored boundary Nonfixed boundaries Ill-defined period Undefined period Embedded period Carry-over ref. period			

(*) From: Lessler & Forsyth (1996: table 11.1, pp. 264-265).

Table 4.5. Condensed Expert Questionnaire Appraisal Coding System

Problems in questionnaire with regard to:		
Comprehension of question	Information processing	Reporting
o Difficult wording => o Unclear wording => o Difficult syntax => o Long question with list of items o Double-barrelled questions o Double-negative questions o Question/answer mismatch o Reference set (perspective) change => o Response task => o …	o Retrieval task => o Long period of recall o Much information needed to answer question o Proxy reporting o Judgement task => o Difficult task (complex calculation, estimation) => o Social desirability o …	o Difficult wording in answering categories => o Unclear wording => o Boundary problems => o Overlapping categories => o Missing categories => o …

=> Indicates a description of problem in question, and suggestions for improvement.

Our scheme should rather serve as a list of questionnaire design risk factors, which may be used by designers of questionnaire prototypes for a practical discussion on hypotheses to be (pre-)tested, and for classifying pre-test findings and discussing possible actions to amend problems."

Expert (re)appraisal is rated rather useful by Campanelli (1997), as based on the conclusions of Presser and Blair (1994): a lot of errors are covered, the method is reliable and cost effective. However, as Campanelli remarks (p. 10): "This may have something to do with the particular experts that were chosen for the study." This indicates that the method is sensitive to the background of the experts, and suggests that several experts (say at least three) have to review the questionnaire. Since the codes are ordered according to the Tourangeau & Rasinki model of the question-and-answer process, this methods reveals errors in the questionnaire for all stages, as is indicated by Sudman et al. (1996). Groves (1996) categorises this method as a special one, revealing problems in the questionnaire without applying it to a respondent.

According to Lessler and Forsyth (1996, p. 260), this scheme is useful in assessing questionnaires and developing plans for questionnaire revision in five ways: "(1) It provides a systematic means for evaluating draft questionnaires. (2) It identifies the question characteristics that effect accuracy. (3) It can be used as a basis for prioritising measurement concerns and developing item revision. (4) It provides a guide for subsequent observation and experimental research. (5) It can be instrumental in building a body of knowledge on question characteristics and associated response errors."

In our view, expert (re)appraisal offers a way to quickly review questions and present the results to clients, covering the question-and-answer process as a whole. The number of problems reported is dependent on the number of experts and their background. A drawback however, is that no respondents participate in this method, making this method not very convincing to customers. Since a formal coding scheme is applied, the method is quantitative in nature, but at the same time systematic. Therefore, the method had been extended with reports of the experts on problematic

parts of the questions and suggestions for improvement. Apart from coming to customer advice, this method also serves to come to hypotheses on the questionnaire to be (pre-) tested with other methods: methods in which the questionnaire is applied to respondents, like focus groups (subsection 4.2.2) and in-depth interviews (subsection 4.2.3). To compare hypotheses and findings, the coding schemes may also be applied in classifying verbal information gathered with these methods.

Table 4.6. An eclectic classification of measurement error risks to assess questionnaires (*)

Problem	Description
Applicability/Suitability	Question is not realistic enough: o Non-existent or inaccessible data o Hypothetical or fictitious data o Data referring to someone else
Question meaning	Question has unclear/ambiguous/unintended meaning: o Misleading or unclear instructions o Reference set (frame) not sufficiently specified o Ambiguity or vagueness in question meaning o Unintended, though univocal, question meaning
Key concept meaning	Key concept: o has unclear or ambiguous meaning o is unknown or unnoticed
Cognitive difficulty	High cognitive burden to respondent: o Difficult recall/recognition o Difficult deduction (estimation, guess) o Difficult judgement o Complicated answer format
Technical difficulty	Complicated question conditions / presentations o Too many key concepts or clauses o Double negation / other syntax complexity o Implicit assumptions o Unclear presentation (lay-out intonation)
Logical flaw	Formulation or routing logically incorrect: o Question-answer (Q/A) discrepancy o Incorrect/incomplete Q/A structure o Conflict with previous answers o Conflict with language rules
Motivation/Affection	Undesirable questions: o Goal of question unclear or insufficient o Too intrusive or personal a question
Social Norms	Unbalanced, directive or non-neutral question: o Non-neutral concept or directive task o Unbalanced or non-neutral answer o Risk of social desirability

(*) From: Akkerboom & Dehue (1997: table 4, pp. 141-142).

4.2.2. Focus groups

Another method for reviewing (draft) questionnaires is the focus group. In an open discussion of 1½ to (at the most) 2 hours, 5 to 10 participants focus on several topics. Generally, the groups are relatively homogeneous, consisting of experts, users (of the data: statisticians, researchers), interviewers or respondents. Preferably, participants don't know each other. Within a (pre-)test program more groups may be moderated. As for focus groups with respondents, each group represents a subgroup of the target population. The focus group is led by a moderator, who uses a list of selected topics as guidance for the discussion. The topics may address the survey as a whole, the questionnaire, the question-and-answer process, or the data collection procedure.

Focus groups are very useful in getting a spectrum of opinions, i.e. an overview of general information. In the discussion, the participants are stimulated to discuss the topics with each other, and to generate new ideas, suggestions and recommendations. The information however, is not very detailed. Also when time is limited, focus groups may be useful. Within one to two weeks, several focus groups may be organised (including recruitment of participants), moderated and analysed, resulting in a broad coverage of a range of topics. There is one exception however: focus groups with employees from businesses. These groups are very difficult to organise, since it is very hard to find employees that are willing to participate and have the time to do so.

To get a focus group started, the following preparations have to be taken care of (SCPR, 1995; Snijkers, 1999; an overview of planning, conducting and analysing focus groups is presented by Hedges, 1985, and more in detail by Stewart and Shamdasani, 1990, and Krueger, 1994):

- *Meeting room.* The meeting room should neither be too small nor too large, preferably with a one-way mirror and a video recording system (for observations). The room should be located and arranged in such a way that the discussion will not be disturbed by others outside the room, and attention will not be distracted from the discussion. In the room a meeting table is needed, that makes a proper arrangement of all participants possible, i.e. all participants should be able to see each other easily (to have a discussion among them), including the moderator (e.g. to make sure that everyone is participating in the discussion).
- *Participants.* Make sure beforehand what kinds of participants are needed on the basis of the research goal. E.g., with regard to a questionnaire that has to be tested, respondents should resemble the target population the questionnaire is applied to. Recruitment of participants with the right background is very important to have a successful discussion. Make sure that in every group the participants are homogeneous, e.g. with regard to age, gender, and occupation. The recommended size of a focus group ranges from 5 to 10 participants. After participants have been selected, they have to be invited, saying where and when they will be expected. To thank participants for their willingness to participate, an incentive is offered to them.
- *Topic list.* The topic list is based on discussions with experts, researchers, and questionnaire designers. Also, the list may be based on results from previous steps in the design process (see

table 4.1), in which e.g. expert appraisals, focus groups or in-depth interviews have been applied.

- *Moderator and observer.* All preparations are carried out by a team of researchers from the Questionnaire Laboratory. They also moderate and observe the focus groups. The observer is not present in the meeting room, but makes notes watching the discussion in an observation room, next to the meeting room, via a one-way mirror or on a video screen. Also, the discussion is audio taped and sometimes video taped.

After the preparations have been taken care of, the focus group is ready to take place (SCPR, 1995; Snijkers, 1999; Molen et al., 1995):

1. *Arrival.* Have about 10 minutes for the participants to arrive, and wait for those who are late. Introduce yourself as the moderator to each participant personally. Make people feel at ease; show interest and respect. People are eager to know what it is all about and what is going to happen. Show people the meeting room. If possible, make a seating arrangement, with those who are dominant, seated next to the moderator, to avoid eye-to-eye contact. Thus, it is less easy for this person to attract the attention of the moderator.
2. *Introduction.* After everybody is seated, the moderator starts with an introduction:
   - Introduce yourself (who you are: name, job) and tell what your role is.
   - Introduce the study, say who it is for and why it is being done.
   - Stress confidentiality and explain how information is used.
   - Say that the discussion is tape (and/or video) recorded, for reasons of analyses afterwards. This makes it necessary for participants to speak one at a time. Do this quickly and efficiently. Agreement to record must be obtained, but don't make a whole discussion out of it. Usually, people forget about the camera after the discussion has started.
   - Say what is expected of the participants: to have an open discussion among themselves.
   - Have the participants introduce themselves: go round.
   - Bear in mind the stages of the group process:
     - Forming: the group is being formed; who is who; going round.
     - Storming: the aim, purpose of the discussion is becoming clear to everyone by initial questioning; what is expected of everyone.
     - Norming: everyone is at ease, relaxed; everything is clear.
     - Performing: this is the productive stage; in this stage the issues on the topic list are being discussed. This is the stage you want to reach, although it could be possible that you don't reach it.
     - Mourning: the group is coming to an end; people are becoming individuals again, although some people might not want the group to end.
3. *Starting the discussion.* After everyone has introduced himself in the introduction (the forming stage), the storming and norming stages are reached:
   - Start the discussion in a non-threatening way. Ask factual questions to establish a good information base and to involve the participants. In this stage it is important that people are made at ease, and are involved in the discussion.

- Have the discussion get started in a smooth way. Do not pursue complicated or sensitive issues too early in the discussion.

4. *During the discussion.* The performing stage has been reached. During this stage the moderator plays an essential role (see table 4.8 for interviewing skills):
   - Stimulate the discussion:
     - Have the participants discuss the topics among themselves. In a good group, people start questioning each other. It is a discussion with everyone joining, in which the moderator is not firing questions, but is the guide.
     - Try to get timid people into the group: “Mr. So-and-so, what do you think about it?”
     - Try to slow dominant people down: “Thank you, but I would like to hear the rest from the group (at the same time raising your hand to him, and looking to the group).”
     - Probe the answers until you are sure that the respondent has replied to the question as fully as possible. Probe long enough to the information you want (but not as deeply as in in-depth interviews). Use neutral, non-directive, follow-up probes (see table 4.7).
     - But, give the respondent time to reply. Do not try to fill in a silence with another question.
     - At the same time, pace the interview in the interest of both the participants and the moderator.
   - Use the group to get:
     - Explanations: “Why do you feel about that in this way?”
     - Ideas, suggestions, and recommendations: “What would you suggest?”
   - Make sure that all issues are covered:
     - Use the topic list.
     - Make a (mental) note of answers or issues that you will need to return to.
     - Go round the participants after each topic has been dealt with.
     - Watch the time.
   - Do not come to the front as a moderator: step aside as a person. It is important that all information comes from the participants: It is not your opinion as a moderator that is of interest, but the opinion of the participants. See table 4.8 for a list of the interviewing skills of the moderator during the discussion.
5. *Ending the discussion:*
   - Try to end the discussion on a positive and completed note. Do not leave issues open. Go round again:
     - Evaluate: “What is your opinion of this discussion?”
     - “What should we do next?”
     - “Do you have any suggestions?”
     - “Is anything left unmentioned?”
   - Do not end the discussion abruptly:
     - Give everyone time to ‘come out’ of the discussion before leaving.
     - Explain again how the information they have given will be used and re-affirm confidentiality.

- At the end of the discussion do not forget to thank the participants for their contribution to the discussion and offer the incentive.

Table 4.7. Follow-up probes

- Non-directive probes to stimulate the respondent to keep talking:
  - The silent probe, in combination with an attentive posture. Show that you are listening by leaning forward (coachman posture) and looking the respondent in the eyes.
  - Humming, nodding, and eye-to-eye contact.
  - Repeating the last words of the respondent in a neutral way (parroting).
  - (after a silence) "Go on, I'm listening."
  - (after a silence) "Tell me what you are thinking?"
- Summarising, checking, and ending an issue. By giving a summary the respondent may be stimulated to keep talking. This is also a way to check whether everything is understood correctly. However, the moderator/interviewer has to be very careful not to put words into the mouth of the respondent. And, in this way an issue may be ended, by asking:
  - "Do we have important aspects left unmentioned?"
  - "Does anyone have anything to add to this issue?"
  - "Do you have anything more to say about this issue?"
- The elaboration probe:
  - "Could you give an example?"
  - "What happened after that?"
- The explanation probe:
  - Repeat the words of the respondent in a questioning way and looking puzzled at him.
  - "I do not understand what you mean. Could you explain that to me?"
  - "Sorry, but I do not follow. What do you mean?"
- The retrospective probe:
  - "Can I take you back to something you said earlier?"
  - "You said …, could I ask you a bit more about that?"
- The meaning-oriented probe:
  - "What do you understand by …?"
  - "What does … mean to you?"
- The 'journalist's' probes: When? Who? Where? What? How?

Table 4.8. Interviewing skills

- Have an open mind. Show interest and respect. Be an attentive listener. Show that you are listening by leaning forward (coachman posture) and looking the respondent in the eyes.
- Look for non-verbal cues, e.g. distress, irritation, and lack of understanding, embarrassment, to probe into.
- Give the respondent time to reply. Step aside as a person. Do not be tempted to fill in a silence, to finish a respondent's answer, or 'to put words into his mouth'.
- Do not divulge personal details about yourself during the interview.
- Do not assume (however obvious an answer seems).
- Do not make comments about the respondent's answers.
- Probe the answers until you are sure that the respondent has replied to the question as fully as possible by using the follow-up probes as listed in table 4.7. Questions and follow-up probes should be clear and open-ended, not closed and not leading.
- Ask one question at a time. Avoid double-barrelled questions.

As for analysing the information, the observation notes are used. On the basis of these notes, the observer and the moderator categorise the information per topic. Preferably, this is done immediately after each focus group. The discussions are audio or video taped as well, as backup for the observer, but also for a more detailed analysis. To analyse the tapes, a matrix is used: the rows represent the topics and the columns the participants. In the cells the comments are transcribed. Also, the tapes may be fully transcribed. However, since this kind of analysis is very laborious and time consuming, reports mostly are based on the observer's notes only. A drawback of only using these notes is that the report may be somewhat subjective.

Focus groups, conducted and analysed in this way, may be used in a more exploratory way in step 1, to gather background information on the survey, but also in a more evaluative way in steps 2 to 5. For instance in pre-testing a questionnaire (in step 2), firstly respondents are invited to complete a self-administered or an interviewer-administered questionnaire, and then have the questionnaire discussed in a focus group. In case of an interviewer-administered questionnaire, a parallel focus group with interviewers could be conducted. (In this situation, information on the question-and-answer process may also be gathered by observations/behavioural coding during the questionnaire administration; see subsection 4.2.4.) In steps 3 to 5, this method may also be applied for respondent and interviewer debriefing (Campanelli, Martin & Rothgeb, 1991; DeMaio & Rothgeb, 1996).

Focus groups are a useful method to quickly get a wide variety of information on issues to be investigated and get new ideas, suggestions, and recommendations. Within a short period of time a lot of respondents are heard. Also, reactions between different groups (in the target population) may be compared. However, it is not possible to research the question-and-answer process in detail. For a detailed evaluation of questions (discussing individual attitudes, issues, wording), in-depth interviews (see the next subsection) are more useful. Therefore, we feel that focus groups used in combination with in-depth interviews increase the value of both methods. Sudman et al. (1996) and Campanelli (1997), in accordance with Fowler & Roman (1992), come to the same conclusion: focus groups and cognitive interviews are complementary, with focus groups being more efficient at identifying definitional problems and in-depth interviews being better at evaluating specific question wording.

4.2.3. In-depth interviews

In the literature a lot of terms are used for in-depth interviews: cognitive interviews (Presser & Blair, 1994; Fowler & Roman, 1992; Campanelli, 1997), concurrent and retrospective protocols (Groves, 1996), expanded interviews (Forsyth & Lessler, 1991), in-depth interviews (Gower, 1991), intensive individual interviews (Belson, 1981; Fowler, 1992, 1995), intensive one-on-one interview (Campanelli, Martin & Rothgeb, 1991), think-aloud interviews (Forsyth & Lessler, 1991; DeMaio & Rothgeb, 1996), think alouds (Sudman et al., 1996). Here we refer to an in-depth

interview as an open interview in which a (self- or interviewer-administered) questionnaire is tested in a laboratory setting with one respondent and an interviewer, in which several techniques are used to provide an in-depth investigation of the question-and-answer process. One of these techniques is thinking aloud in combination with follow-up probes; other techniques that will be discussed here are meaning-oriented probing, paraphrasing, targeted test questions and vignettes.

Like all methods discussed in this chapter, the purpose of in-depth interviews is to get information on difficulties in the questionnaire, in order to improve it. At the Questionnaire Laboratory of Statistics Netherlands in-depth interviews have been used to identify difficulties in question comprehension, perceptions of response tasks, memory recall strategies, interpretations of question reference periods, difficulties in selecting a response, and reactions to sensitive questions (see also table 4.6). This method is generally used in cognitive laboratories and has proven to be useful in identifying these difficulties, as is also concluded by Campanelli (1997) in her review.

To get started with in-depth interviews about the same preparations have to be taken as with focus groups (see subsection 4.2.2):

- *Interviewing room.* An interviewing room has to be organised with a table, making a 90° arrangement possible. The respondent and the interviewer should not sit opposite to each other, to make it possible for the respondent to look away while thinking. The room should have a one-way mirror or a video system for making observations possible, e.g. by an observer, other researchers or customers, but without potential distraction.
- *Number of interviews and recruiting respondents.* The number of interviews (5 to 50) has to be chosen, depending on the wishes of the customer and time constraints. Then, volunteering respondents with background characteristics that resemble the target population have to be recruited. Recruitment of respondents can be done in several ways: by asking colleagues and respondents who had been invited before, whether they know relatives, friends, neighbours, etc., who will be willing to participate in interviews; by making a call in the internal newspaper and in regional home-to-home papers for voluntary respondents; by addressing clubs; etc. These respondents are registered in a data base, with their name, address, etc. including information on availability, and times of participation. To make sure that they will not become ‘professional respondents’, they are invited three times at the most, but preferably not more than twice. Also some background variables, like gender, date of birth, education level, occupation, are included in the data base, in order to select respondents with specific characteristics. Respondents to participate in in-depth interviews (or focus groups), then, are selected from this data base. After selection, they have to be invited, saying where and when they will be expected. To thank respondents for their willingness to participate, an incentive is offered to them.
- *Interviewing protocol.* A team of researchers design the (pre-)test protocol on the basis of step 1 (table 4.1), or the monitoring of standardised interviews in other steps, in which potential problems with questions have been formulated. For each question that has to be tested, the protocol describes how the question-and-answer process will be investigated. The questionnaire

and the laboratory techniques are integrated into one (pre-) test protocol (Snijkers, 1997; chapter 5). Usually, the questionnaire is too long to (pre-)test all questions, so a selection of 30 to 60 questions has to be made, with in-depth interviews lasting 1½ hours at the most. The number of questions that can be tested depends on the techniques that are being used. E.g., thinking aloud is a very time consuming technique. So, when a lot of thinking aloud is done, the maximum number of questions is limited to say 30. After the protocol has been designed, it has to be tested. The protocol is tested within the team (being both interviewers and respondents) and other colleagues as respondents, to find out whether the protocol works as intended and to train the interviewers in getting to know and using the protocol. During this testing stage the protocol may be improved.

- *Interviewer and observer*. A team of researchers from the Questionnaire Laboratory takes care of all preparations. They also conduct and observe the in-depth interviews. The observer is not present in the interviewing room, but makes notes of the interview from behind a one-way mirror or on a video screen. The interviews are recorded on audio tape and sometimes on video tape.

After the preparations have been completed, the in-depth interviews can be conducted (Campanelli, 1993; SCPR, 1995; Snijkers, 1999; Molen et al., 1995):

1. *Arrival.* After the respondent has arrived, introduce yourself, as the interviewer, and have the respondent seated where you would like him to sit. Make the respondent feel at ease; show interest and respect. The respondent is eager to know what it is all about and what is going to happen. Show the respondent where to sit.
2. *Introduction.* Then, the interviewer starts the actual interview with an introduction:
   - Introduce yourself once again as interviewer and tell what your role is.
   - Introduce the study, say who it is for and why it is being done.
   - Stress confidentiality and explain how information is used.
   - Say that the interview is tape (or video) recorded, for reasons of analyses afterwards. Do this quickly and efficiently. Agreement to record must be obtained, but don't make a whole discussion out of it. Usually, people forget the microphone after the interview has started.
   - Say what is expected of the respondent and give instructions on what he has to do. This depends on the technique that is applied during the in-depth interview. It's not a test of the respondent but of the questionnaire.
3. *Starting the interview.* After all formalities have been taken care of, the actual interview is about to start:
   - Start the interview by asking a few example questions that are not used for analysis. In this way the interview is started in a non-threatening way.
   - Interviewer and respondent roles are established through conduct of these questions. These questions are used for a natural warm-up.
   - Do not pursue complicated or sensitive issues too early in the interview.

4. *During the interview*. One could say that at this point, like in focus groups, the performing stage has been reached. During this stage the interviewer plays an essential role (see table 4.8 for interviewing skills):
   - Stimulate the respondent:
     - Probe the answers until you are sure that the respondent has replied to the question as fully as possible. Probe long and deeply enough to the information you want. Use neutral, non-directive, follow-up probes (see table 4.7).
     - But, give the respondent time to reply. Do not try to fill in a silence with another question.
     - At the same time, pace the interview in the interest of both the respondent and the interviewer.
   - Make sure that all issues are covered:
     - Use the pre-test protocol (Snijkers, 1997; chapter 5).
     - Make a (mental) note of answers or issues that you will need to return to.
     - Watch the time.
   - Do not come to the front as an interviewer: step aside as a person. It is important that all information comes from the respondent: It is not your opinion as an interviewer that is of interest, but the opinion of the respondent. See table 4.8 for a list of the interviewing skills.
5. *Ending the interview:*
   - Try to end the interview on a positive and completed note. Do not leave issues open:
     - Evaluate: “What do you think of think of this interview?”
     - “What should we do next?”
     - “Do you have any suggestions?”
     - “Is anything left unmentioned?”
   - Do not end the interview abruptly:
     - Give the respondent time to ‘come out’ of the interview before leaving.
     - Explain again how the information will be used and re-affirm confidentiality.
     - At the end of the interview thank the respondent for his contribution to the research and offer the incentive.

As for analysing the interviews, the audio tapes are used. When time is short however, the notes of the observers are used, with the tapes as backup. Preferably, immediately after each interview, the interviewer and observer discuss the notes and come to conclusions on the interview. When there is more time, or the client wants a detailed report, the tapes are analysed using the codes in table 4.6 or the form in figure 4.1 (Snijkers et al., 1994). The tapes are not transcribed as a whole, because that would take too much time. Sometimes the interviews are also videotape recorded. This is done to make a compilation of respondent reactions to be shown to the customer, in addition to a report.

The form in figure 4.1 combines several techniques for analysing the tapes: transcripts, expert appraisal (see table 4.5), and observations/behavioural coding (see subsection 4.2.4, table 4.11).

Figure 4.1. Analysis form for in-depth interviews

Respondent number : <number> Analysis scheme QLAB: <survey>
Interviewer : <name>
Coder : <name>
Qlab technique : o thinking aloud o test question o <other technique>

<Question text>

REMARKS FROM RESPONDENT

Answer:

o R expresses doubts between categories
o answer doesn't fit in list of categories
o R changes answer
o R doesn't know / never thought about / refuses

o calculated
o estimated
o guessed

Remarks from respondent on confidence in own answer:
…

Task:

o easy
o so - so / not too difficult
o difficult

Respondent evaluation of question:

o R says to have no problem with Q
o R requests for repetition of Q
o R requests spontaneously for clarification of Q =>
o R interrupts question reading =>

Respondent has problems with

Comprehension of question:

o Difficult wording =>
o Unclear wording =>
o Difficult syntax =>
o Long question, list of items
o Double-barrelled questions
o Double-negative questions
o Question/answer mismatch
o Reference set change =>
o Response task =>
o …

Information processing

o Retrieval task =>
o Long period of recall
o Much information needed to answer question
o Proxy reporting
o Judgement task =>
o Difficult task (complex calculation, estimation) =>
o Social desirability
o …

Reporting

o Difficult wording in answering categories =>
o Unclear wording =>
o Boundary problems =>
o Overlapping categories =>
o Missing categories =>
o …

CODER INTERPRETATION OF RESPONSE

o immediate answer
o (long) silence after question reading
o breaks of silence while answering
o doubt between categories
o doesn't know / never thought about / refusal
o …

For every interview and for every question a form is filled out by at least 2 researchers/interviewers from the team. The interviews are analysed by more than 1 researcher in order to get intersubjective interpretations of the question-and-answer process. Typically, an interviewer analyses his own interviews. The appropriate codes are ticked and relevant comments of the

respondent are transcribed. Also, the researcher may give his interpretation of what is going on. Then, after all questions in all interviews have been analysed, the results are summarised and combined into an overall form for every question. If necessary, e.g. in case of discrepancies in interpretations, the researchers discuss what is going on and come to a final interpretation. Those summary forms are the basis for a report.

As mentioned before, several techniques can be used within in-depth interviews. A technique that is generally applied is *thinking aloud*. With this technique the respondent is asked to report his thoughts while answering a survey question (Snijkers, 1997; chapter 5). This may be done concurrently or retrospectively. Concurrently thinking aloud (Forsyth & Lessler, 1991; DeMaio & Rothgeb, 1996) means that the respondent is prompted to think aloud while answering a question. At the beginning of the interview respondents are instructed to think aloud (using a few warm-up questions), and are stimulated by the interviewer during are interview to keep talking by the use of neutral, non-directive, *follow-up probes* (see table 4.7; Snijkers, 1997; chapter 5). Mostly, respondents need to be prompted, but sometimes respondents keep talking by themselves. With retrospective think alouds (as used by Belson, 1981; and described by Forsyth and Lessler, 1991) a standardised interview is conducted, followed by an in-depth interview in which the respondent is asked to tell his thoughts while the question had been answered. To help the respondent recall his thoughts, a video tape of the interview may be reviewed during the think-aloud session.

Sudman et al. (1996) describe another way of applying retrospective think alouds. They have respondents first answer a question and then immediately afterwards probe into the question-and-answer process. According to them “it is unrealistic to assume that respondents can remember how they retrieved information for one question after proceeding with additional questions.” (p. 35) This is a major drawback of the retrospective method as used by Belson (1981). As for concurrent think alouds, Sudman et al. state that respondents firstly should report their thoughts and then answer the question. However, the distinction between concurrent and retrospective think alouds as proposed by Sudman et al. is quite theoretical. In practice, we see that the ordering of answering a question and thinking aloud differs among respondents. As Sudman et al. (1996) put it (p. 34): “Some respondents, usually those with higher levels of education and greater verbal facility, find concurrently thinking aloud an easy and interesting process and give rich protocols. Others, however, need prompting, turning what is intended to be a concurrent think aloud into a retrospective think aloud. That is, some respondents answer the question very quickly but do not report their thought process. (…) It should be noted that some respondents who provide unprompted verbal protocols do so after answering the question, thus muddying the distinction between concurrent and retrospective think alouds.”

At Statistics Netherlands we have chosen to apply concurrent think alouds (Snijkers et al., 1994), following the guidelines for organising and conducting in-depth interviews as described above. According to us, this technique allows the respondent to react spontaneously, making the interview appear like an ordinary conversation. By most respondents this method is easily

understood, although for less-educated respondents reporting of their thoughts might be a tough job. Furthermore, concurrently thinking aloud does not lean heavily on the memory of the respondent and is easier to organise (no extra equipment like video screens are needed and the respondent is less long occupied), as is the case with retrospective think alouds. A drawback of concurrent think alouds is that it breaks up the flow of the standardised interview: the pace of the interview slows down and the question-and-answer process may be disturbed. This means that the respondent might react differently to a question in a standardised interview than in a concurrent think-aloud session.

Thinking aloud is a general technique in which all aspects of the question-and-answer process may be addressed. This technique is not directed on specific aspects of the process, and results in a lot of information. This technique is time-consuming, and may not be very efficient. When the respondent is talking, information may be gathered that is not of direct interest to the researcher. Particular aspects of the question or the question-and-answer process may be focused on by targeted techniques.

One of these techniques is *meaning-oriented probing* (see table 4.7; Gerber & Wellens, 1997; Snijkers, 1995a, 1995b, 1997; chapters 5, 6 and 7). This probe is used to investigate how the respondent interprets a word or a term in a question, e.g. by asking: “What do you understand by ‘net income’ in this question?” This techniques focuses on question comprehension. It is very frequently used, without or within thinking aloud, and results in very useful information. In combination with thinking aloud, it may happen that the respondent already has given his view on a term, which makes this probe unnecessary to be asked.

Another targeted technique that is directed on question comprehension is *paraphrasing* (Lessler & Forsyth, 1991; DeMaio & Rothgeb, 1996). In paraphrasing, the respondent is requested to repeat the question in his own words. This technique is used to determine whether respondents understand a question as intended, as with meaning-oriented probing, but also whether the response task is interpreted correctly. Paraphrasing could be used to find out how respondents memorise the question, what words are difficult to them (e.g. jargon) and which are not, and to get suggestions for rewording the question in the words of the respondent. However, this task appeared to be very difficult for respondents (Snijkers et al., 1994). First of all, it is not easy to make clear to respondents what their task is, i.e. what they have to do. When the respondent was asked how he would state the question to his mother or neighbour, the task became clear. However, they still did not understand why they had to do this. Secondly, once the task was clear, respondents had difficulties in finding other words on the spot. They replied that they would not state the question differently. This makes this technique not very informative. As a result this technique has not been used very frequently.

A technique for investigating particular aspects of a question is the *targeted test question, evaluation question, or elaboration probe*. In the questionnaire focused questions are embedded in

order to investigate how e.g. a specific question is interpreted or how an answer is obtained. Also, a question may be rephrased at the end of a questionnaire, to get additional information on the original question. For instance, when one wants to probe into the number of rooms in a house, an elaboration probe: "Did you include the kitchen, the bathroom and/or the cellar in the number of rooms?" could be asked (Snijkers, 1995a; Snijkers et al., 1999; chapters 6 and 8). In combination with *follow-up probes*, this technique results in additional and detailed information on the issues of interest. However, this technique can only be applied to a few questions without unduly lengthening the interview. Furthermore, like with think alouds, the use of targeted test questions and follow-up probes within the interview may slow down the interview and influence responses to subsequent questions. Targeted questions at the end of the questionnaire do not have the latter disadvantage. Apart from its use in in-depth interviews, this technique may also be used in field interviews (in steps 3 to 5, table 4.1), using closed questions to monitor the quality of survey questions.

The last targeted technique to be used within an in-depth interview described here, is classification of *vignettes* (Forsyth & Lessler, 1991; DeMaio & Rothgeb, 1996). Vignettes are short descriptions of situations. Respondents are asked to select category labels that best describe that situation according to their interpretation of the concept. For example, a respondent may be given a vignette on vocational training ('Robert is studying for aircraft constructor. He went to Fokker last year for a 6-months apprenticeship.'; Snijkers, 1995b; chapter 7), and is expected to tell whether that situation should be categorised as vocational training, training or something else. Preferably, this technique is applied at the end of an interview, not to interfere with the survey questions. The goal of this technique is to determine how respondents interpret the topics of interest, i.e. to investigate where respondent's definitions differ from not-given, official definitions. With this technique the concepts of interest are defined from the viewpoint of the respondents. Thus, like with 'vocational training', vignettes indicate what activities may be underreported. This technique may also be used in field interviews or in re-interviews (in steps 3 and 4, table 4.1; Campanelli, Martin & Rothgeb, 1991; DeMaio & Rothgeb, 1996), after a categorical system of interpretations has been developed in step 2. We feel however, that this technique is not very informative on the question-and-answer process. The information gathered is quantitative in nature. With vignettes we know that respondents interpret situations differently, but we don't know why. Vignettes in combination with meaning-oriented and follow-up probes are a better way to get information on comprehension.

As said before, in-depth interviews are used very frequently in cognitive laboratories to pre-test questionnaires (in step 2). Our experience, in accordance with Groves (1996) and Campanelli (1997), is that this method offers a way to investigate all aspects of the question-and-answer process in detail, resulting in a lot of information, depending on the technique(s) used. Not only problems with questions are detected, but also ideas about how to improve the questions are obtained. Within one pre-test program, the questions may be changed accordingly, in order to test the rephrased questions in the next in-depth interviews. However, this method is laborious and time

consuming. And sometimes even too much information is gathered, information that is not of direct interest. This holds especially for thinking aloud in combination with follow-up probing. Less laborious and time consuming are in-depth interviews using only meaning-oriented probes. When information has to be gathered quickly, the focus group may be the better method.

Another limitation of this method, as well as the focus group method, is that the sample size is quite small, therefore resulting in unrepresentative conclusions (Fowler & Cannell, 1996). Also, the fact that these methods are conducted in laboratory conditions may be seen as a drawback. Thus, the resulting information of in-depth interviews is qualitative in nature. This makes it hard to convince customers (statisticians) of the usefulness of these methods. Our experience however, is that showing video tapes of respondents trying to answer a difficult question to customers is quite convincing. A quantitative method for monitoring questionnaires, that may used in combination with focus groups and in-depth interviews, is behavioural coding.

4.2.4. Observations of interviewer and respondent behaviour

The last method that will be described here is behavioural or reaction coding. With this method observers code the behaviour or reactions of the interviewer and the respondent during a standardised survey interview, i.e. code the question-and-answer process. Coders (interviewers, experts, and researchers) observe the interview from behind a one-way mirror or watch a video screen, and systematically code the behaviour of the interviewer and the respondent. Apart from these systematic observations that will be described in this subsection, we have already mentioned unstructured observations with focus groups and in-depth interviews, in which an observer notes all aspects of respondent behaviour that may be of interest in the study.

This method may be used to analyse interviews that precede focus groups or in-depth interviews, but also to monitor standardised field interviews, like telephone or recorded face-to-face interviews. Oksenberg, Cannell and Kalton (1991) described this method in detail. According to them, this method in used in the final stage of the development process. They state (p. 349): “No matter how extensive the development work in questionnaire construction is, however, there remains the need to test the resulting questionnaire under field conditions before it is finally adopted for the actual survey. It is this pre-test stage that is the subject of the research reported here.” The stage Oksenberg and his colleagues refer to are steps 3 and 4 in the 5-step (pre-)test model (table 4.1).

According to Oksenberg et al. (1991, pp. 349-350) the method is directed at identifying questions that are troublesome to interviewers and respondents: Interviewers may have difficulties reading the questions (because of complex sentence structure or words that are difficult to pronounce), and respondents may face difficulties with regard to comprehending the question (difficult vocabulary, complex sentence, difficult task), a lack of common understanding (terms and

concepts are understood differently by different respondents, or are not interpreted as the researcher intended), and cognitive processing of information (because they are not able or not willing to make the effort needed to provide an adequate answer, or because the information needed to answer adequately is inaccessible to them). These goals are also stated by Forsyth and Lessler (1991, p. 396): "This procedure is directed at identifying questions that are difficult to ask or to understand, lack of common understanding of terms and concepts, and difficulties in processing the requested information." Or as Fowler and Cannell (1996, p. 35) put it: This method focuses largely on "cognitive problems, like question comprehension, concept clarity, and response task difficulty." According to Sudman et al. (1996) and Groves (1996) this method primarily is directed at comprehension of the question. "Coders look for two indications of problems with a question: respondents make remarks asking for clarification, and respondent's answer to the question is inadequate." (Sudman et al., 1996, pp. 22-23.) Groves (1996) also points out that this method is directed at the interaction of the interviewer, questionnaire and respondent.

The codes Oksenberg et al. (1991: figure 1, p. 352) mention, are listed in table 4.9. Fowler and Cannell (1996: table 2.4, p. 29) also present these codes. This list of diagnostic codes is a minimal list that has been developed over a number of studies for identifying questions that are problematic for interviewers or respondents.

Table 4.9. Interviewer and respondent behaviour codes (*)

Interviewer question reading codes:		Definition
E	• Exact	• Interviewer reads the question exactly as printed.
S	• Slight change*	• Interviewer reads the question changing a minor word that does not alter question meaning.
M	• Major change*	• Interviewer changes the question such that the meaning is altered. Interviewer does not complete reading the question.
Respondent behaviour codes:		**Definition**
1	• Interruption with answer*	• Respondent interrupts initial question reading with answer.
2	• Clarification*	• Respondent asks for clarification of question, or makes statement indicating uncertainty about question meaning.
3	• Adequate answer	• Respondent gives answer that meets question objectives.
4	• Qualified answer*	• Respondent gives answer that meets question objectives, but is qualified to indicate uncertainty about accuracy.
5	• Inadequate answer*	• Respondent gives answer that does not meet question objective.
6	• Don't know*	• Respondent gives a "don't know" or equivalent answer.
7	• Refusal to answer*	• Respondent refuses to answer the question.

* Indicates a potential problem with the question.
(*) From: Oksenberg et al. (1991: figure 1, p. 352).

In table 4.9 some codes are marked, indicating that these codes show a potential problem with the question. However, a major problem with behaviour coding is that the codes do not identify the source of the problem it covers. The codes show that there is a problem, but do not uncover the actual problem. In addition to the codes, Fowler and Cannell (1996: table 2.5, p. 32) describe a number of problems that go together with the codes. These problems are listed in table 4.10.

Table 4.10. Common question problems affecting interviewer and respondent behaviour (*)

Behaviour	Common problem
Interviewer misreads question	• Awkward wording (difficult to pronounce). • Introductions that interviewers find superfluous or do not fit context. • Dangling clauses at end of question. • Missing explanation of response task.
Respondent requests clarification	• Unclear, undefined terms. • Unclear response task. • Poor question order; respondents cannot recall response alternatives.
Respondent interrupts	• Wording results in a complete question before the question is finished. • Respondent does not realise response alternatives will be given.
Respondent gives inadequate answer or interviewer probes	• Meaning of question is unclear. • Unclear response task or poor question order so respondent cannot recall response alternatives. • Response task does not fit answers. • Response task is difficult, requires effort to recall or asks too much detail.

(*) From: Fowler & Cannell (1996: table 2.5, p. 32).

Table 4.11. Codes for respondent reaction

Respondent reaction	Stage of question-and-answer process and problems covered
	Stage 1: Interpretation or comprehension stage
1. Respondent requests for repetition of question 2. Respondent requests for clarification of question 3. Respondent interrupts question reading	The first two codes indicate that the question may be hard to understand, due to unclear or undefined terms, long question, complex syntax, poor question order (influencing the interpretation), unclear response task, or respondents cannot recall response alternatives. The third code indicates that apparently there is no interpretation problem, or the question may be too long and redundant, or the respondent doesn't realise that response options will be given.
	Stages 3 and 4: Judgement and reporting stage
4. Respondent expresses doubt between response alternatives 5. Respondent gives answer that doesn't fit in the list of response alternatives	These two codes have to do with the list of response alternatives: it may be hard for the respondent to make up his mind and choose for only one option (the alternatives may not be clearly mutually exclusive), or response options may be missing.
	Stages 2 and 3: Retrieval and judgement stage
6. Respondent changes answer	The sixth code indicates that it may be hard for respondents to recall the correct information.
7. Respondent gets irritated by question	The last code expresses a general feeling of irritation by the respondent towards the question.

The Questionnaire Laboratory at Statistics Netherlands developed an alternative set of codes for respondent reactions, based on the codes introduced by Oksenburg et al. (1991; see table 4.9). We felt that some of these codes could not be associated to observable respondent or interviewer behaviour. Furthermore, these codes were subject to discussion: When is a change in question reading a slight change? And, When does a coder know that an answer is qualified? According to us, the data resulting form these codes would not be very reliable. Therefore, we developed a set of codes associated as much as possible with observable respondent behaviour in a one-on-one relation. This set was to be used for monitoring CAPI and CATI questions during the course of an interview by interviewers (Snijkers et al., 1999; chapter 8). Since the codes had to be used by

interviewers, we only defined a limited number of respondent codes. These codes were chosen in such a way to cover all four stages of the question-and-answer process. The codes and the problems they identify, are listed in table 4.11.

Campanelli (1997) rates behavioural coding rather low with regard to usefulness, because it is not considered to be a sensitive enough tool to identify the exact source of the problem it uncovers, e.g. the term that is problematic, or the response alternatives that are not exclusive. Furthermore it does not identify as many problems as the other methods. These limitations are also mentioned by Oksenberg et al. (1991). On the other hand this method gives quick results, and offers a quantitative, replicable and systematic measure for diagnosing potential question problems. Therefore, we feel that behavioural coding is useful to quickly identifying troublesome questions. In this way, the method can be used in steps 3 to 5 to monitor standardised field interviews, as was also pointed out by Oksenberg et al. (1991, see above). These questions should then be tested cognitively, to get more insight in why the problems occur. Thus, this method gives input to step 2 by formulating hypotheses on questions, that are to be investigated in more detail with methods like focus groups or in-depth interviews.

4.3. Conclusions

The methods described in detail in section 4.2, have been applied successfully at the Questionnaire Laboratory at Statistics Netherlands to pre-test questionnaires (step 2 in table 4.1). The methods most frequently used are expert (re)appraisal, in-depth interviews (with thinking aloud and follow-up, meaning-oriented, and elaboration probing), and focus groups. Less frequently used are paraphrasing and classification of vignettes. Expert (re)appraisal is used when time constrains are very strict. When there is more time to execute the research, say 2 to 4 weeks, interviewers and respondents are invited to participate in focus groups. And in case we are given even more time (1 to 4 months) to carry out a study in which respondents had to be heard, in-depth interviews are conducted.

In carrying out a pre-test program with application of these methods, we try to meet the following scientific standards, to guarantee quality of the research:

- The methods are executed in a standardised way, meaning that all interviewers, observers and researchers carry out their jobs in one and the same way. One way to do this is to train interviewers and observers in their task during the preparation phase of a study. Other ways to get standardisation are the use of the computer for carrying out in-depth interviews (Computer-Assisted Qualitative Interviewing; Snijkers, 1997; chapter 5), the use of one topic list in focus groups, and the use of standardised tools for analysis. Thus, interviewer and observer variance will be reduced. And furthermore, with standardisation, procedures are repeatable and results are reproducible (Tucker, 1997). In this way we try to get reliable data and conclusions of the study.

- Usually two or (preferably) three interviewers and observers participate in a study. As for expert (re)appraisal at least three experts report on the questionnaire. Focus groups and in-depth interviews are prepared, carried out and analysed within the team. And, together they come to final conclusions. Thus, interviewer and observer bias will be reduced. In this way we try to get inter-subjective or more objective conclusions.
- Also we try to apply triangulation. Preferably, in one study two to three methods are being applied. In this way, as many problems as possible are revealed, but also problems uncovered with one method may be confirmed with other methods. Thus, the information gathered is both confirmatory and complementary. The combinations we find very useful are:
  - In case of a not yet fully designed questionnaire: expert appraisal and focus groups. Problems, uncovered with expert (re)appraisal, are studied in more detailed with focus groups. When the newly designed questionnaire is ready for application, the focus groups may be followed by in-depth interviews. With this method, the question-and-answer process is investigated in even more detailed.
  - When a questionnaire has been designed fully or has to be redesigned, a useful combination of methods is respondent reaction coding and focus groups or in-depth interviews. Like in the first combination, problems uncovered with reaction coding may be studied in more detail with the other methods. And since reaction coding is a quantitative method, the limitations of focus groups and in-depth interviews (a small, unrepresentative sample and the unique conditions under which the interviews occur) are compensated for. As for in-depth interviews, several techniques may be applied as well: thinking aloud and follow-up probes may be combined with targeted test questions.

  Thus, the validity of the conclusions is controlled for.

In practice however, the carefulness of the execution of a study depends on time, budget and staff constraints. So, the quality criteria mentioned above are not always met. Although, carrying out a study within time and budget limits are also quality issues.

As we have seen, these methods not only can be used in pre-testing questionnaires but also in other steps to development a data collection procedure. In section 4.1, these methods have been presented within the context of the 5-step (pre-)test model for data collection development (table 4.1). In a full pre-test program all aspects of the survey are carefully tested in advance, following this model and the guidelines presented in this chapter.

Table 4.1 indicates for each step what methods can be used, the topics that are addressed, the recommended test size, and the general results. A short definition of these methods is presented in table 4.2. Other overviews (Sudman et al., 1996; Groves, 1996) indicate the aspects of the question-and-answer process (comprehension of wording and interpretation of response task, information retrieval, judgement, and reporting) the methods are targeted at. Forsyth and Lessler (1991) categorise the methods according the type of method. Campanelli (1997) rates the methods on usefulness, i.e. number of problems uncovered.

Table 4.12. Features of cognitive laboratory methods

Method	Step(s) in 5-step model	Question-and-answer process	Type of method *	Resulting information		Costs in application		
				qualitative / quantitative	usefulness **	laboratory staff ***	time to complete	(pre-) test size
Expert (re)appraisal	1 – 3	All stages	expert evaluation	quantitative, extended with qualitative reports	++: depending on the number and experience of the experts	3 – 10 cognitive experts	1 week	– (questionnaire not applied to respondents)
Focus group	1 – 5	All stages	group method	qualitative	+++: overview of general information, new ideas	1 assistant 1 – 3 researchers	2 – 4 weeks	small scale
In-depth interview:	2	depending on technique:	expanded interview	qualitative	depending on technique:	1 assistant 2 – 3 researchers	1 – 4 mnths	small scale
Thinking aloud and Follow-up probing	2	All stages	expanded interview	qualitative	++++: detailed information			small scale
Meaning-oriented probing	2	comprehension of wording	targeted method	qualitative	+++: detailed information			small scale
Paraphrasing	2	comprehension of wording and task	targeted method; difficult task for respondent	qualitative	+: no or little additional information			small scale
Targeted test question	2 – 5	All stages, dependent of focus of question	targeted method	qualitative and quantitative	+++: detailed information (with follow-up probes)			small and large scale
Vignettes	2 – 4	comprehension of wording	targeted method	quantitative	+: not very informative on Q-and-A process			small and large scale
Observations Behavioural / reaction coding	2 – 5	All stages, but accent on comprehension of wording and task, reporting	expert evaluation	quantitative	+: diagnosing potential question problems; no information on the source	1 assistant 2 – 3 researchers	1 – 6 mnths	small and large scale

* Taxonomy as presented by Forsyth & Lessler (1991)
** Usefulness: number of problems covered (Campanelli, 1997)
*** Assistant: recruitment, organisation, observing; researchers: moderating, observing, analyses

To conclude this chapter, the features of the cognitive laboratory methods are summarised in table 4.12, according to our experiences at Statistics Netherlands, as described above and in section 4.2. In the first column the methods is listed. The second column refers to the step(s) in the 5-step model in which the method can be applied. The third column refers to the stages in the question-and-answer process the method reveals information on. In column four, the methods are categorised according to the overview by Forsyth & Lessler (1991). In the next two columns the resulting information is indicated: qualitative (verbal reports) or quantitative information in column 5, and the usefulness of the information (the number of problems uncovered) in column 6. In the last three columns the costs in applying the methods are summarised: number of laboratory staff involved, time needed to execute the method (including preparations, execution, analyses and reporting), and test size (detailed information on the test size is presented in table 4.1).

The method that is most effective, in number of problems detected, is in-depth interviewing in combination with thinking aloud, follow-up probing, meaning-oriented probing, and targeted test questions. This method, however, may take some time to complete (1 to 3 months). Focus groups are less time consuming (2-4 weeks), and are also very effective. The number of problems detected with expert appraisal depends on the expertise of the experts. In general, this method is less effective than in-depth interviewing and focus groups, since the questionnaire is not applied to respondents. An advantage of this method is that it can be applied very quickly (with results within 1 week). Behavioural coding is not very effective too, since it diagnoses only potential question problems, and provides no information on the source of the problems. The information gathered with this method is of quantitative nature, which makes it useful in combination with qualitative methods like in-depth interviews and focus groups. However, it takes some time and effort to get enough data. Paraphrasing and vignettes, to be used in in-depth interviews (and, as for vignettes, also in field interviews), are not effective. Therefore, these techniques are not used very often.

Examples of studies in which these methods have been used, are presented in the next chapters:

- In the chapters 6 and 7, laboratory studies using in-depth interviews are described. In chapter 6 a questionnaire on income is pre-tested, using thinking aloud with follow-up probes, meaning-oriented probing and targeted test questions (Snijkers, 1994, 1995a). Chapter 7 describes a pre-test on questions on training from the European Community Household Panel, using thinking aloud with follow-up probes, meaning-oriented probing and vignettes (Snijkers, 1995b; Van de Donk & Snijkers, 1995).
- In chapter 8 a study on POLS (Continuous Survey on Living Conditions) is described (Snijkers et al., 1999). In this study all steps of the 5-step pre-test model have been followed. In this chapter we will concentrate on step 3: the qualitative operational test, using respondent reaction coding and elaboration probing in field interviews. In the preceding laboratory study (in step 2) in-depth interviews and focus groups have been conducted.
- In chapter 9 an example of a focus group study is presented within the context of a model to stimulate respondent to participate in the survey (Snijkers & Luppes, 2000). With this study a business survey is pre-tested: the redesigned Annual Establishment Production Survey. While

in the other chapters questionnaires for households and individuals are tested, here we have an example of applying cognitive methods to business surveys.

In all studies, except for chapter 9, the CAQI method (Computer-Assisted Qualitative Interviewing) was used for applying cognitive methods. The CAQI method will be described in chapter 5 (Snijkers, 1997).

References

Akkerboom, H., and F. Dehue, 1997, The Dutch Model of Data Collection Development for Official Surveys. (International Journal of Public Opinion Research, Vol. 9, No. 2, pp. 126-145.)

Akkerboom, H., F. Dehue, and G. Snijkers, 1998, Qualitative Tests for Data Collection (Re-) Design. The Case of POLS. (In: Bakker & Winkels (eds.), Integration of Household Surveys: Design, Advantage and Methods, pp. 20-25.)

Akkerboom, H., and A. Luiten, 1996, Selecting Pre-testing Tools according to a Model of Questionnaire Development, with Illustrations concerning Patient Satisfaction with medical care. (In: Proceedings of the Conference of the American Association for Public Opinion Research, AAPOR, Salt Lake City.)

Bakker, B., and J. Winkels (eds.), 1998, Integration of Household Surveys: Design, Advantage and Methods. (Netherlands Official Statistics, Special Issue, Vol. 13, Summer 1998. Statistics Netherlands, Voorburg/Heerlen.)

Banks, R., C. Christie, J. Currall, J. Francis, P. Harris, B. Lee, J. Martin, C. Payne, and A. Westlake (eds.), 1999, … Leading Survey and Statistical Computing into the New Millennium. Proceedings of the Third ASC International Conference. (Association for Survey Computing, Chesham, Bucks, UK.)

Belson, W.A., 1981, The Design and Understanding of Survey Questions. (Gower Publishing Company, London.)

Biemer, P.P., R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (eds.), 1991, Measurement Errors in Surveys. (Wiley, New York.)

Campanelli, P., 1993, A 1-day Seminar on Cognitive Interviewing. (Statistics Netherlands, Department of Statistical Methods, April 16, Heerlen.)

Campanelli, P., 1997, Testing Survey Questions: New Directions in Cognitive Interviewing. (Bulletin de Methodologie Sociologique, No. 55, June 1997, pp. 5-17.)

Campanelli, P., E. Martin, and J. Rothgeb, 1991, The Use of Respondent and Interviewer Debriefing Studies as a Way to study Response Error in Survey Data. (The Statistician, Vol. 40, pp. 253-264.)

Campanelli, P., J. Rothgeb, J. Esposito, and A. Polivska, 1991, Methodologies for Evaluating Survey Questions: An Illustration from a CPS CATI/RDD Test. (Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ.)

Cook, Th.D., and D.T. Campbell, 1979, Quasi-Experimentation. Design & Analysis Issues for Field Settings. (Rand M^{c}Nally College Publishing Company, Chicago.)

Dehue, F., 1997, Standard Protocols for Developing and Testing Data Collection Procedures. (In Dutch: Standaard protocollen voor de ontwikkeling en de test van dataverzamelings-procedures. CBS report, BPA No. H7247-97-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

DeMaio, Th., and J. Rothgeb, 1996, Cognitive Interviewing Techniques In the Lab and in the Field. (In: Schwarz & Sudman (eds.), Answering Questions, pp. 177-195.)

Esposito, J., P. Campanelli, J. Rothgeb and A. Polivska, 1991, Determining which Questions are best: Methodologies for Evaluating Survey Questions. (In: American Statistical Association, Proceedings of the Section on Survey Research Methods, Alexandria, VA.)

Foddy, 1993, Constructing Questions for Interviews and Questionnaires. Theory and Practice in Social Research. (Cambridge University Press, Cambridge.)

Forsyth, B.H., and J.T. Lessler, 1991, Cognitive Laboratory Methods: A Taxonomy. (In: Biemer et al. (eds.), Measurement Errors in Surveys, pp. 393-418.)

Forsyth. B.H., J.T. Lessler and M.L. Hubbard, 1992, Cognitive Evaluation of the Questionnaire. (In: Turner et al., Survey Measurement of Drug Use: Methodological Studies.)

Fowler, F.J., 1992, Developing Procedures for Testing Questions in Government Surveys. (Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, Fl.)

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

Fowler, F.J., and Ch.F. Cannell, 1996, Using Behavioral Coding to Identify Cognitive Problems with Survey Questions. (In: Schwarz & Sudman (eds.), Answering Questions, pp. 15-36.)

Fowler, F.J., and A.M. Roman, 1992, A Study of Approaches to Survey Question Evaluation. Final Report for U.S. Bureau of the Census. (Center for Survey Research, Boston.)

Gerber, E., and T. Wellens, 1997, Perspectives on Pre-testing: 'Cognition' in the Cognitive Interview. (Bulletin de Methodologie Sociologique, No. 55, June, pp. 18-39.)

Gower, A., 1991, The Questionnaire Design Resource Centre's Role in Questionnaire Research and development at Statistics Canada. (Paper presented at the 48th Session of the International Statistical Institute, September 1991, Caïro, Egypt.)

Gower, A., B. Bélanger and A. Paletta, 1998, The Contribution of Qualitative Methods in Improving the Quality of Survey Data. (Paper presented at the "Quality Criteria in Survey Research" Seminar, June 1998, Cadenabbia, Italy.)

Groves, R.M., 1996, How do we know what we think they think is really what they think? (In: Schwarz & Sudman (eds.), Answering Questions, pp. 389-402.)

Hedges, A., 1985, Group Interviewing. (In: Walker (ed.), 1985, pp. 71-91.)

Hox, J.J., 1997, From Theoretical Concept to Survey Question. (In: Lyberg et al. (eds.), Survey Measurement and Process Quality, pp. 47-69.)

Krueger, R.A., 1994, Focus Groups. A Practical Guide for Applied Research. (Sage, London.)

Lessler, J.T., and B.H. Forsyth, 1996, A Coding System for Appraising Questionnaires. (In: Schwarz & Sudman (eds.), Answering Questions, pp. 259-291.)

Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, D. Trewin (eds.), 1997, Survey Measurement and Process Quality. (Wiley, New York.)

Molen, H.T. van der, F. Kluytmans, and M. Kramer, 1995, Having discussions: Skills and Models. (In Dutch: Gespreksvoering: Vaardigheden en modellen. Wolters-Noordhof, Groningen / Open Universiteit, Heerlen.)

Oksenberg, L., Ch. Cannell, and G. Kalton, 1991, New Strategies for Pretesting Survey Questions. (Journal of Official Statistics, Vol. 7, No. 3, pp. 349-365.)

Presser, S., and J. Blair, 1994, Survey Pretesting: Do different Methods produce different Results? (Sociological Methodology, Vol. 24, pp. 73-104.)

Saris, W.E., 1998, The split-ballot MTMM experiment: An alternative way to evaluate the quality of questions. (Research paper, University of Amsterdam, Amsterdam.)

Schwarz, N., 1997, Questionnaire Design: The Rocky Road from Concepts to Answers. (In: Lyberg et al. (eds.), Survey Measurement and Process Quality, pp. 29-45.)

Schwarz, N., and S. Sudman (eds.), 1996, Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. (Jossey-Bass, San Francisco.)

SCPR (Social & Community Planning Research), 1995, A Short Course in Focus Group Techniques (as used in Qualitative Research). (Statistics Netherlands, Department of Data Collection Methodology, March 9-10, Heerlen.)

Snijkers, G., 1994, What is the total net household income? Cognitive interviews on income questions. Design and Results. (In Dutch: 'Wat is het totale netto huishoudinkomen? Vragenlab-onderzoek naar de gecoördineerde verkorte inkomensvraag. Opzet en resultaten. Internal, unpublished CBS report. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1995a, What is the total net household income? Cognitive interviews on income questions. (Paper presented at the International Conference on Survey Measurement and process Quality, April 1-4, Bristol, UK.)

Snijkers, G., 1995b, Pre-tests on ECHP Questions on Daily Activities, Pension Schemes, and Education and Training: Final Results. (Internal CBS report, BPA no. H8170-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1997, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. (Bulletin de Methodologie Sociologique, No. 55, pp. 93-107.)

Snijkers, G., 1999, Guidelines for Internal Quality Audits: Background, Execution, and Skills. (In Dutch: Handleiding Kwaliteitsaudits bij Divisie GVZ: Achtergrond, uitvoering en vaardigheden. Internal CBS report, BPA no. H712-99-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., H. Akkerboom, and J. Nollen-Dijcks, 1994, 'What do we ask the voter?' Probing into Questions on the Election Survey. The Research Design. (In dutch: 'Wat vragen we de kiezer?' Doorvragen naar vragen uit het kiezersonderzoek. De onderzoeksopzet.' Internal CBS report, BPA no. 114-94-M1. Statistics Netherlands, Department of Statistical Methods, Heerlen.)

Snijkers, G., E. de Leeuw, D. Hoezen, and I. Kuijpers, 1999, Computer-Assisted Qualitative Interviewing: Testing and Quality Assessment of CAPI and CATI Questionnaires in the field. (In: Banks et al. (eds.), … Leading Survey and Statistical Computing into the New Millennium, pp. 231-258.)

Snijkers, G., and M. Luppes, 2000, The Best of Two Worlds: Total Design Method and New Kontiv Design. An operational Model to improve Respondent Co-operation. (Paper presented at the Second International Conference on Establishment Surveys (ICES-II): Survey Methods for Businesses, Farms, and Institutions, June 17-21, Buffalo, New York.) (In: American Statistical Association, Proceedings of Invited Paper, pp. 361-371. Alexandria, Virginia.)

Stewart, D.W., and P.N. Shamdasani, 1990, Focus Groups. Theory and Practice. (Applied Social Research Methods Series, Vol. 20, Sage, London.)

Sudman, S., N.M. Bradburn, and N. Schwarz, 1996, Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. (Jossey-Bass, San Francisco.)

Tourangeau, R., and K.A. Rasinksi, 1988, Cognitive Processes underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol. 103, No. 3, pp. 299-314.)

Tucker, C., 1997, Methodological Issues surrounding the Application of Cognitive Psychology in Survey Research. (Bulletin de Methodologie Sociologique, No. 55, pp. 67-92.)

Turner, C.F., J.T. Lessler and J.C. Gfroerer (eds.), 1992, Survey Measurement of Drug Use: Methodological Studies. (DHHS Publications No. 92-1929, U.S. Government Printing Office, Washington D.C.)

USCB (U.S. Census Bureau), 1998, Pretesting Policy and Options: Demographic Surveys at the Census Bureau. (U.S. Department of Commerce, Washington, DC.)

Van de Donk, P., and G. Snijkers, 1995, Testing on ECHP-questions on daily activities, pension schemes, and education and training: final report. (CBS report, BPA no. H2876-95-SIP. Statistics Netherlands, Division of Socio-Economic Statistics, Department of Integration and Presentation, Heerlen.)

Van den Brakel, 2001, Design and Analysis of Experiments Embedded in Complex Sample Surveys. (Ph.D. thesis, Erasmus University, Rotterdam.)

Walker, R. (ed.), 1985, Applied Qualitative Research. (Dartmouth, Aldershot.)

Chapter 5
Methods 2

Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires

Summary: *At Statistics Netherlands, almost all household surveys are computer-assisted. Within this setting, the Questionnaire Laboratory at Statistics Netherlands developed a computer-assisted method for pre-testing computerised questionnaires using the cognitive laboratory methods (in particular in-depth interviews) as discussed in the last chapter. We called this method 'Computer-Assisted Qualitative Interviewing' (CAQI). In this chapter the CAQI method for conducting cognitive interviews will be discussed. A CAQI protocol is expressed by instruction screens and probes built around the questions that are to be tested. CAQI has been applied in several pre-test studies both in the laboratory and in the field (of which examples will be presented in the following chapters). Our experience is that CAQI creates realistic fieldwork conditions in the laboratory, and helps to conduct cognitive interviews in a standardised way. CAQI is a workable method, resulting in improved qualitative information on the question-and-answer process*

Keywords: *Pre-testing Questionnaires, Cognitive Interviews, Computer-Assisted Interviewing.*

- Paper presented at the Fourth International Social Science Methodology Conference, July 1-5, 1996, University of Essex, Colchester, UK.
- Paper published in: Bulletin de Methodologie Sociologique, June 1997, No. 55, pp. 93-107.
- The CAQI method as discussed in this paper, has been developed in co-operation with Hans Akkerboom. The author is grateful to Marko Roos for his contributions to section 4 of this paper.

5.1. Introduction

The Questionnaire Laboratory at Statistics Netherlands started its work in 1992. This Laboratory offers facilities and methods for designing and pre-testing questionnaires. A pre-test program typically consists of a small number of cognitive interviews (say up to 50). In these interviews respondents are requested to provide information on the question-and-answer process by using cognitive interviewing techniques like thinking aloud, paraphrasing, meaning-oriented probing, vignettes, etc.

At Statistics Netherlands, almost all household surveys are carried out by means of computer-assisted personal or telephone interviewing (CAPI or CATI) using Blaise. Within this setting (see chapters 2 and 3), we designed a computer-assisted method for pre-testing computerised questionnaires in a cognitive laboratory. This computer-assisted method for pre-testing computerised questionnaires is meant to provide *qualitative* information about how questions are processed and how answers are obtained. Thus, the CAI interview becomes a CAQI interview. By adding cognitive interviewing techniques to a standardised CAI interview, the original Blaise questionnaire is extended with a 'CAQI protocol'.

CAQI is applied in many pre-test programs, e.g. in pre-testing questions from the European Community Household Panel (ECHP) questionnaire (Snijkers, 1995; chapter 7). In this chapter CAQI will be explained using the ECHP pre-test program as example. How CAQI works and what it looks like will be described in section 5.2. In section 5.3 our reasons for applying CAQI and our experiences with CAQI will be described. To see whether CAQI really works according to our experiences, the ECHP pre-test interviews are coded with regard to interviewer and respondent behaviour. The results of the behavioural coding will be presented in section 5.4. In this chapter the ECHP pre-test interviews is chosen as an example because some of these pre-test interviews were conducted with paper-and-pencil. This gave us the opportunity to compare paper interviewing with CAQI. This comparison is also described in section 5.4. Conclusions will be discussed in section 5.5.

5.2. Computer-Assisted Qualitative Interviewing (CAQI)

The starting point for CAQI is a computer-assisted questionnaire (in our case a Blaise questionnaire). To get a CAQI protocol the cognitive interviewing protocol is integrated in this questionnaire. The CAQI protocol is expressed by instruction screens built around the (group of) questions that are to be pre-tested. These screens tell the interviewer (1) by which instructions and examples the respondent should learn his cognitive interviewing tasks, (2) when to insert which cognitive interviewing techniques during the standardised interview (headings like "Thinking aloud: arbitrary order of thinking aloud and answering"), and (3) how to apply these techniques, say by neutral probes to stimulate the respondent (e.g. "What are you thinking now?"). This means

that all instructions are in the computer. For conducting a CAQI interview the interviewer starts the interviewing program on the computer and every single step (both with regard to the standardised and the cognitive interviewing protocol) is put forward automatically.

In the ECHP pre-test program the main technique used was thinking aloud. This technique was used to find out what general problems respondents have with understanding, retrieving information and answering the questions. In the ECHP pre-test interviews, this technique was applied with almost every question. Another technique used was meaning-oriented probing, to get the interpretation of specific terms in the questions.

In this section we will describe what CAQI looks like, with regard to the CAQI introduction (in subsection 5.2.1), thinking aloud (5.2.2) and meaning-oriented probing (5.2.3).

5.2.1. CAQI introduction

First of all, in the CAQI protocol for pre-testing the ECHP questions the Blaise questionnaire was extended with instructions for the respondent. These were a general introduction of the interview and an instruction on thinking aloud. After the CAQI program was started (and a respondent id-number was entered into the computer), these instruction screens were the first to appear on the computer screen.

In figure 5.1 and 5.2 these instruction screens are shown. The text within the lines, prefaced with '>> LAB:', is instruction merely for the cognitive interviewer. This text should not be read aloud to the respondent, but tells the interviewer what cognitive interviewing technique(s) will be applied. These interviewer instructions are reminders for the interviewer, telling what is coming up and which probes can be used. The other lines have to be read to the respondent. It is not necessary to read these lines verbatim, as long as the message is the same. What is important is that the message is clear to the respondent. To verify this, on a separate screen a question is posed to the respondent asking whether the intention of this interview is clear (figure 5.3).

All together, this is a somewhat lengthy introduction, but it is important to tell the respondent what is expected of him and to make him feel at ease. In case thinking aloud is not the main technique used in the cognitive interviews, the instructions on thinking aloud (figure 5.2a and 5.2b) are not necessary.

Figure 5.1. General introduction to the respondent of the interview

```
>> LAB: Aim of this interview.

Good morning/afternoon,

We have invited you to participate in this interview.
This is not an ordinary interview. Right now, I am going to ask you a
number of questions.
The aim of this interview is to find out whether these questions are
clear to PEOPLE LIKE YOU.
Whenever a question is not clear to you, please say so.
With the help of your COMMENTS, we try to change the questions, so that
they will be clear afterwards.
For that purpose, we will also carry out such an interview with
A FEW OTHER PEOPLE.

>> LAB: Press <ENTER> to continue.
```

Figure 5.2a. Instruction thinking aloud for the respondent

```
>> LAB: Instruction THINKING ALOUD.

With the following questions, I would like to hear from you how you react
to the question, i.e. YOU FOR YOUR PART.
With each of the following questions, please THINK ALOUD and tell me
everything that occurs to you. This means that you tell me ALL that you
are THINKING.
Hence you MENTION EVERYTHING that comes to your mind.
Please tell me also what you think while looking for an answer.
It is up to you to decide when you give the answer.
What really MATTERS is that you THINK ALOUD.

>> LAB: Press <ENTER> to continue.
```

Figure 5.2b. Instruction thinking aloud for the interviewer

```
>> LAB: Think aloud guidelines:
        DO NOT ASSUME,
        BUT encourage respondents to keep thinking aloud,
            by using NEUTRAL (non-directive) PROBES.

>> LAB: Question may be repeated.
>>      NEUTRAL (non-directive) PROBES ONLY:
           - Go on, I'm listening!
           - What are you thinking now?
           - What comes to your mind?
           - Why did you choose this answer?
           - Can you tell me a bit more about that?
           - I would like to hear your opinion about that!
           - When?  Who?  Where?  What?  How?
           - THE SILENT PROBE!
>>      DO NOT ASSUME.

>> LAB: Press <ENTER> to continue.
```

Figure 5.3. Question for respondent

```
>> LAB: Lab-question.

 Is clear to you what the intention is?

>> LAB: Press <ENTER> to continue.
```

5.2.2. CAQI think-aloud question

After the introduction, the actual interview begins. The difference with a standardised interview is that now instructions for the interviewer dealing with the cognitive interview protocol are also shown on the screen.

In an ordinary standardised interview a question on the screen might look like the question in figure 5.4, taken from the ECHP questionnaire. In CAQI however, a screen for a think-aloud question looks like the screen in figure 5.5. We see that instructions with regard to the cognitive interviewing technique to be used (in this case thinking aloud), have been built around the survey question. At the top of the screen an interviewer instruction says that this a think-aloud question. The interviewer now knows what to do: read the question verbatim to the respondent and stimulate him to think aloud. At the bottom of the screen examples of probes that can be used are displayed. These probes are reminders for the interviewer, he may use. However, the interviewer may use other probes as well, whenever he thinks these are necessary. In this way, the interviewer can concentrate fully on the respondent's answer, without worrying about what to say and what to do.

Figure 5.4. Standardised question

```
Have you at any time since January 1994 been in vocational education or
training, including any part-time or short courses?

(enter code)
1: Yes
2: No
```

Figure 5.5. CAQI screen for thinking-aloud question

```
>> LAB: THINK-ALOUD question:
        Arbitrary order of thinking aloud and answering.

 Have you at any time since January 1994 been in VOCATIONAL
 EDUCATION or TRAINING, including any part-time or short courses?

>> LAB: Question may be repeated.
>>      NEUTRAL (non-directive) PROBES ONLY:
             - Go on, I'm listening!
             - What are you thinking now?
             - What comes to your mind?
             - Why did you choose this answer?
             - Can you tell me a bit more about that?
             - I would like to hear your opinion about that!
             - When?  Who?  Where?  What?  How?
             - THE SILENT PROBE!
>>      DO NOT ASSUME.
```

5.2.3. CAQI meaning-oriented probe

In the question in figure 5.5 we see that some words are printed in capitals. This indicates that the interviewer has to probe into the meaning of these words. It could very well be that respondents spontaneously give an interpretation of these words. In that case it is not necessary for the interviewer to go into the meaning of these words. But otherwise the interviewer has to give a meaning-oriented probe. As a reminder for the interviewer to do this, the next screen (figure 5.6) comes up. (In the ECHP-CAQI questionnaire figure 5.6 immediately came after figure 5.5.) The line at the top of the screen indicates that this is a meaning-oriented probe (and not an ordinary survey question). At the bottom of the screen the question is reprinted, because in Blaise on a single screen only separate questions are displayed. So now all information is on one screen and the interviewer does not have to page up.

In the ECHP-CAQI questionnaire the next screen (the one that came right after figure 5.6) was a screen that looked like figure 5.6, but now with a meaning-oriented probe for ‘training’. Thus, like with the thinking-aloud question, the computer takes care of the next probe (or question).

Figure 5.6. CAQI screen for meaning-oriented probing

```
>> LAB: MEANING-ORIENTED PROBE.

 What do you understand by VOCATIONAL EDUCATION in this question?

>> LAB: Question may be repeated

        The question was:

 Have you at any time since January 1994 been in VOCATIONAL
 EDUCATION or TRAINING, including any part-time or short courses?


>> LAB: Press <ENTER> to continue.
```

5.3. Experiences with CAQI

The staff of the Questionnaire Laboratory (who conduct the cognitive interviews) has a lot of experience on how CAQI works out in practice. In this section our experiences with CAQI will be described (subsection 5.3.2), as well the reasons for applying CAQI (subsection 5.3.1).

5.3.1. Reasons for applying CAQI

In a cognitive interview using CAQI a number of participants play a role. These are: the interviewer, the respondent, the Blaise questionnaire program, the cognitive interviewing protocol and the computer. As we have seen in section 5.2, in a CAQI interview the interviewer poses questions and probes to a respondent, using a computer, in which the cognitive protocol is embedded in the Blaise questionnaire. To get a perfect interview setting the respondent should be able to react spontaneously to all questions, without being disturbed in any way. The interviewer should know what to do: how to handle the Blaise questionnaire, how to handle the computer, and when and how to apply what cognitive interviewing techniques.

As for the general setting, an important reason for applying CAQI at Statistics Netherlands is that in the CAQI interviews the same questionnaire is used as in the field, using all properties of CAI like complex skip patterns, editing answers, personalization of question wording, etc. (Snijkers, 1992). Thus, a general property of CAQI is that cognitive interviews are more or less structured, i.e. with a standardised protocol all cognitive interviews are conducted in the same way (i.e. no parts of the cognitive protocol are skipped) and reflect fieldwork conditions.

Another important property of CAQI is that the computer helps the interviewer to control the flow of the interview, i.e. applying the appropriate cognitive technique, using the right probes, and still leaving enough freedom to the interviewer to apply other probes whenever necessary. Because the protocol is embedded in the computer interviewers can do the interviews almost right away, without much instruction. This may be very handy when interviewers have to replace others at the last minute (e.g. when they have fallen ill). And furthermore, because of the automated flow and the embedded instructions, the computer helps the interviewer to focus on the communication with the respondent.

5.3.2. Experiences with CAQI

Our experience is that, since the protocol is embedded in the computer, the interviewer does not have to worry about what to do (i.e. applying the appropriate techniques and probes). However, among our staff there was a differing meaning on the information that had to be put on the screen. Novice interviewers remarked that they felt very comfortable during the interview knowing that

everything was on the screen: they could always fall back on the probes listed on the screen. More experienced interviewers however noted that the screen was too full with text: they did not have the time to read all that was printed on the screen during an interview. For them it was enough to put only reminders on the screen saying what technique to apply.

Although the computer is of help to the interviewer in some ways, it may also become a disturbing factor during the interview. This may happen when the interviewer does not know how to handle the computer, or when something unexpected happens on the screen. Then the attention of the interviewer may be distracted from the respondent. Therefore, it is important that the interviewer knows how to handle the computer, the Blaise questionnaire and the CAQI protocol. This makes it necessary to have the interviewers fully trained in handling the computer, and using Blaise. Also a general training in applying cognitive interviewing techniques is required. And for every individual CAQI questionnaire detailed instructions are necessary. When interviewers know how to handle the computer they may be able to do a CAQI interview right away, however they feel that they should know why a CAQI protocol is like it is. Then they know what information to probe for.

In subsection 5.3.1 we have said that the computer helps the interviewer to control the flow of the interview. However, to get a fully elaborated CAQI protocol, every step has to be carefully planned in advance. By writing a CAQI questionnaire program (like with ordinary standardised CAI), the designer is urged to prepare all parts of the cognitive interview carefully. Also, the CAQI questionnaire has to be tested carefully. This is done in test interviews with colleagues, in which the interviewers learn the new CAQI questionnaire, and also changes can be made to the protocol. Altogether the design of a CAQI protocol may take some time, depending on the questionnaire to be pre-tested and the complexity of the cognitive protocol. And besides, an important condition to use CAQI is that the researcher should be able to program or have the cognitive protocol programmed in the source text of the survey questionnaire.

As for the respondent our experience is that they are not hindered by the computer. They can express themselves freely. Sometimes we get remarks indicating that the use of the computer gives a professional status to the whole setting. One thing is important though: the physical setting. The computer should not be like a block in between the interviewer and the respondent. This can be avoided with the 90-degrees setting.

So, with regard to CAQI we may state that:

1. With CAQI the cognitive protocol is conducted in a standardised way, i.e. as it should be, without omissions: no meaning-oriented probes skipped and adequate probing to get the adequate information on the question-answer process. Thus, the computer helps the interviewer to control the flow of the interview.
2. CAQI leaves enough freedom for the respondent to react spontaneously.
3. CAQI helps the interviewer to focus on the communication with the respondent.

5.4. Interviewer and respondent behaviour in CAQI interviews

To see whether our feelings concerning CAQI really hold, we analysed the ECHP pre-test interviews (Snijkers, 1995, see chapter 7) with regard to interviewer and respondent behaviour.

In the ECHP pre-test program (European Community Household Panel) about 20 questions on present work and daily activities, pension schemes, education and training were pre-tested. The cognitive interviewing techniques used were thinking aloud and meaning-oriented probing. The ECHP questions were pre-tested in two phases. A preliminary phase was carried out with 8 respondents to get some first results quickly, and to develop the cognitive interviewing protocol in detail. These cognitive interviews were conducted by one interviewer with a paper version of the questionnaire and the cognitive interviewing protocol. The following phase consisted of CAQI interviews with 24 respondents carried out by 4 interviewers, one of whom was the interviewer who did the interviews in the preliminary phase.

The CAQI interviews of three interviewers have been analysed with regard to interviewer and respondent behaviour. Seven interviews, conducted by the author, were not coded because he had programmed the cognitive interviewing protocol in Blaise. So, he knew exactly what to do, which means that his interviews are incomparable to the others with regard to interviewer behaviour. Although the numbers are very small, the design of the ECHP pre-test program gives us the possibility to evaluate how CAQI works in practice (as compared to a paper-and-pencil pre-test program) and test our statements (see subsection 5.3.2).

As for the interviewer behaviour we looked for omissions in the execution of the cognitive interviewing protocol: number of meaning-oriented probes skipped, no adequate probing (i.e. limited probing or no probing, resulting in limited or no adequate information on the question-answer process), and interviewer reactions indicating that he was not paying attention to the respondent or the protocol. As for the probing behaviour we looked for interviews in which for at least three times the probing was skipped or not adequate. We have chosen for three omissions in an interview, because occasionally, probes may be skipped e.g. because the interviewer simply forgets to probe. But, when this happens more than a few times in an interview it becomes a serious omission in the execution of the protocol. Interviewer reactions are coded whenever they appeared.

As for the respondent behaviour we looked for thinking-aloud behaviour: in how many interviews did the respondent think aloud, i.e. spontaneous thinking aloud (concurrent thinking aloud, see Sudman et al., 1996) or thinking aloud after probing (retrospective thinking aloud). To make sure that thinking aloud did not occur by chance we counted interviews in which thinking aloud occurred with at least 3 questions.

The results of the coding are presented in table 5.1. The numbers in table 5.1 give a qualitative impression of interviewer and respondent behaviour in the two modes in the ECHP pre-test interviews.

Table 5.1. Interviewer and respondent behaviour in ECHP pre-test interviews

Interviewer and respondent behaviour	Paper	CAQI		
	int.1	int.1	int.2	int.3
Number of interviews conducted	8	13	3	1
Interviewer behaviour:				
- meaning-oriented probes skipped (≥ 3 Qs)	4	0	0	0
- no adequate probing (≥ 3 Qs)	8	13	2	0
- interviewer reactions	6	3	3	1
Respondent behaviour:				
- concurrent thinking aloud (≥ 3 Qs)	5	9	3	1
- retrospective thinking aloud (≥ 3 Qs)	8	13	3	1

With regard to the respondent behaviour, table 5.1 indicates that concurrent thinking aloud occurred in 5 out of the 8 paper-mode interviews with 3 or more questions conducted by interviewer 1. In CAQI this occurred in 9 out of 13 interviews done by interviewer 1. Retrospective thinking aloud occurred in all interviews for all modes for all interviewers. These numbers indicate that with regard to thinking aloud, the behaviour of the respondent may not be influenced by the mode. The thinking-aloud behaviour may be influenced however by respondent characteristics, interviewer characteristics, and the instructions at the beginning of the interview.

However, the mode is of influence to the execution of the cognitive protocol: in 4 out of the 8 paper interviews with 3 or more questions, probes were unintentionally skipped. In the CAQI interviews this did not occur at all. This indicates that the chance to skip a meaning-oriented probe with a paper protocol is higher than in CAQI.

The interviewer behaviour ‘no adequate probing’ occurred in the paper version as well as in the CAQI interviews in the same proportions. This indicates that adequate probing is not dependent on the mode. Adequate probing is a specific interviewer skill: the ability to decide whether the respondent has provided enough information on a question or to continue probing.

The interviewer reactions, indicating that the interviewer was not paying attention to the respondent or the protocol, that occurred during the paper interviews all had to do with the skipping pattern: the interviewer had to find out what the next question is. Some of the reactions that occurred were:

- “H’m. ... Let me see. ... What is the next question?”
- “No, wait a minute. Now I am making a mistake. I’m not supposed to ask you these questions. I have to skip these ones.”

In the paper protocol the probes were listed in between the survey questions (see subsection 5.5.2). Because of this the interviewer lost track of the skipping pattern.

The interviewer reactions during the CAQI interviews all had to do with handling the computer and the Blaise program. The following reactions occurred:

- "The computer is acting strangely. I don't know what is going on."
- "Whenever the computer is beeping, I am making a mistake."
  "Sorry, it's beeping again. That's my mistake."
- The interviewer forgot to enter an answer code and says: "What is going on?"
- "I don't know how to enter 'don't know'. Do you mind when I enter 'no'."

5.5. Conclusions and discussion

5.5.1. Conclusions

The numbers in table 5.1 underline our experiences as stated in subsection 5.3.2 with CAQI with regard to skipping meaning-oriented probes by interviewers and the think-aloud behaviour of respondents: in CAQI, meaning-oriented probes (displayed on individual screens) are less likely to be skipped (than in paper interviewing) and the respondent has enough freedom to react spontaneously.

However, adequate probing is not influenced by the mode (paper or CAQI). With CAQI, interviewers may quickly learn how to conduct a CAQI interview when they know how to handle the computer and the Blaise questionnaire, without much instruction. But, there is more to adequate probing then only doing what is on the screen: adequate probing is dependent on the skills of the interviewer and whether interviewers know what information to probe for. This means that detailed instructions with regard to the why and how of a CAQI protocol are required. Then, interviewers are able to probe for the right information.

With regard to focussing on the communication with the respondent, CAQI does not always help the interviewer. As we have seen, handling of the computer and the Blaise questionnaire did sometimes cause a slight or a general breakdown of the interview. This means that the interviewer is not giving any attention to the respondent, but is trying to get the computer in order. To avoid these breakdowns it is important that the interviewers know how to handle the computer and the Blaise questionnaire.

So, now we may conclude that:

1. With CAQI, the cognitive protocol is conducted as it should be, without omissions. In CAQI no probes that are on individual screens are being skipped (which is the case in paper interviewing). In this way CAQI helps the interviewer to control the flow of the interview, as long as the protocol is correctly programmed. However, inadequate probing may still occur.
2. The respondent is not hindered by the computer in his reactions.

3. CAQI helps the interviewer to focus on the communication with the respondent, as long as the interviewer knows how to handle the computer and the Blaise questionnaire.

All together, we may conclude that with CAQI the quality of the information on the question-answer process resulting from cognitive interviews is of higher standards, as compared to paper-and-pencil interviewing.

5.5.2. Discussion

Now, we have discussed the CAQI method, but would it be possible to test computerised questionnaire in another way. These other methods may be: (1) using a paper version of the questionnaire in which the cognitive interviewing protocol is integrated or (2) using the computer for the standardised questionnaire together with a paper version of the cognitive interviewing protocol. However, these methods do not seem very practicable to us.

As for the first method, due to complex skipping patterns, it is almost always impossible to get a paper version of a CAI questionnaire. However, this method may work in case the skipping pattern of the questionnaire is very simple. An important drawback of this method however, is that the interviewer has to take care of the routing herself. This may be difficult to do, since the skipping pattern (although straightforward) may not be very transparent due to the cognitive protocol that is interwoven in the questionnaire. E.g., routing instructions say that with a particular answer a question has to be skipped, but before going to the next question a few probes have to be put forward. While going into these probes, the interviewer has to remember the skipping pattern. This means that the attention is drawn away from the respondent.

This method was used in the first phase of the ECHP pre-test program. As the numbers in table 5.1 indicate, with this method it is very likely that omissions occur in the execution of the cognitive protocol.

The second method is: using the computer for the standardised questionnaire together with a paper version of the cognitive interviewing protocol. With this method the interviewer has to integrate the questionnaire and the cognitive interviewing protocol herself during the interview. This means that the interviewer has to pay attention to two instruments: the interviewer has to look at the computer screen, then on the paper for the cognitive interviewing protocol, and above all has to pay attention to the respondent. This method was used in pre-testing the questionnaire of the Continuous Survey on Living Conditions (Dehue, 1996). Our experience is that in this way the interviewer is busy to get the protocol right, which leads to omissions in the cognitive interview: probes are unintentionally skipped.

All together, our experience is that CAQI is a workable method to pre-test computerised questionnaires, resulting in better information on the question-answer process. However, an important condition to use CAQI is that in practice it should be possible to have the cognitive protocol integrated in the source text of the survey questionnaire, i.e. have the facilities and the time. As for the execution of the CAQI interviews, the interviewers should be well trained in standardised as well as cognitive interviewing techniques, and in the use of the computer as well as the Blaise program. Furthermore, before they start doing the cognitive interviews they should be instructed in the goals of the pre-test program.

As mentioned in the introduction of this chapter, the CAQI method has been applied in many pre-test studies at the Questionnaire Laboratory. In the next chapters examples of these studies are presented. In chapters 6 and 7 laboratory studies will be discussed, respectively on income questions and the ECHP questions. In the study that will be described in chapter 8 CAQI has been applied in the field, in order to investigate questions from the Continuous Survey on Living Conditions. In this chapter the experiences with CAQI in the field will be discussed (section 8.2).

References

Dehue, 1996, Pre-testing the POLS questionnaire (in Dutch). In: W. de Heer, 1996, POLS: Kwalitatief testtraject. CBS report: H4299-96-GWM. (Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1992, Computer-Assisted Interviewing: Telephone or Personal ? A Literature Study. (In: A. Westlake et al. (eds.), 1992, Survey and Statistical Computing, pp. 137-146.)

Snijkers, G, 1995, Pre-tests on ECHP questions on daily activities, pension schemes, and education and training: final results. CBS report: H8170-95-GWM. (Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Sudman, S., N.M. Bradburn and N. Schwarz, 1996, Thinking about answers. The application of Cognitive Processes to Survey Methodology. (Jossey-Bass Publishers, San Francisco.)

Westlake, A., R. Banks, C. Payne and T. Orchard (eds.), 1992, Survey and Statistical Computing. (North-Holland, Elsevier Science Publishers, Amsterdam.)

Chapter 6
Case Study 1: Testing Income Questions

What is the Total Net Household Income? Cognitive Interviews on Income Questions

Summary: *In 1994 the cognitive laboratory of Statistics Netherlands pre-tested a newly designed, small set of questions on net household income, with regard to the question-and-answer process. This set of questions was meant to be a standardised and co-ordinated set to be used in every survey in which income is asked for as a background variable. In this chapter the design of the pre-test study is described, as well as identified problems and suggestions for improving the questions. In the study CAQI in-depth interviews have been conducted using thinking aloud with follow-up probes, meaning-oriented probing and targeted test questions.*
This pre-test study was carried out by order of the former Department of Social Accounts of Statistics Netherlands. This chapter is a report that has been written (Snijkers, 1994) to present the results to the client. In discussions preceding the study, the client indicated that they were mainly interested in a detailed description of identified problems in the questions. But, the client was also interested in suggestions for improvement, to be followed by a discussion on implementation of these recommendations. Unfortunately, due to a major reorganisation of Statistics Netherlands in 1994, our client no longer existed at the end of 1994. This complicated the presentation of the results. This chapter provides a first example of a laboratory pre-test study (step 2 of the 5-step (pre-)test model of data collection development) and how its results (i.e. identified problems and recommendations) are presented to the client.

Keywords: *Pre-testing Questionnaires, Cognitive Interviews, Net Household Income.*

- Paper presented at the International Conference on Survey Measurement and Process Quality, April 1-4, 1995, Bristol, UK.

6.1. Introduction

On request of the former Department of Social Accounts of Statistics Netherlands, in 1994 the cognitive laboratory of Statistics Netherlands has pre-tested a newly designed, set of questions on income. This set of questions was supposed to measure the total of all earnings received on the postal or bank account, for all income sources and for all persons who share their daily life (as members of a household) in a predefined period of time. The data had to be used to cluster households with regard to income. Furthermore, this set was intended to be a *standardised and co-ordinated CBS* question, meaning that in every CBS-household survey (both CAPI and CATI) in which income is asked for as a background variable, this set has to be used.

Since the questions had to be used on the phone as well, some additional demands to the design of the questions were put forward. The set of questions should be small, and individual questions should not be too long. And furthermore, the respondents should be able to answer the questions easily, on top of their heads, without having to look for the necessary data in their personal archives or having to ask other household members. In the pre-test, these aspects were investigated, with regard to the question-and-answer process (interpretation, retrieval, evaluation of information, and response selection; see Tourangeau & Rasinski, 1988).

With these aspects in mind, researchers of the department of Social Accounts had designed a small set of questions, consisting of five questions that deal with the following aspects (see appendix 6.1 for the question wordings):

- the head breadwinner: the person in the household with the highest net income;
- the major income source of the head breadwinner: the breadwinner's highest income source;
- the total net household income, in an ordinary month (for employed), or in the last twelve months (for self-employed or employed who don't know their monthly income);
- paying health-insurance premium;
- income-tax restitution.

The key question is the question on the total net household income. This question is worded in three different ways, depending on the major income source of the head breadwinner. Respondents coming from households, in which the head breadwinner gets wages, salary, pension or a benefit, are asked to report the total net *monthly* household income. Respondents, who don't know this income, are asked in addition to report the total net *yearly* household income. Respondents coming from households in which the highest income is earned in their own business (self-employed) are also asked to report their total net *yearly* household income (although this question is worded differently). The two questions on the head breadwinner serve as routing questions to get to the right key questions. The last two questions should indicate whether the income should be corrected for health insurance and income-tax effects.

In this chapter, the cognitive laboratory study with regard to this set of questions will be described (Snijkers, 1994, 1995). In section 6.3 the conclusions will be presented, as well as suggestions for improvement. In section 6.4 the results will be described in detail. But first of all, the design of the study will be described in the next section.

6.2. Design of the pre-test study

In the period of 5-25 April 1994 the set of questions was pre-tested. The aim of the pre-test was to investigate the question-and-answer process: find out how people react to these questions, what the difficulties in the questions are with respect to the mental process, why those difficulties occur, and how the questions can be improved. All questions should be tested in the appropriate order, but the focus should be on the key question. Apart from this general investigation on difficulties with regard to the question-and-answer process, a specific research question was posed: Is it possible for respondents to answer the questions on top of their head?

The questions were pre-tested with 31 face-to-face in-depth interviews, using the CAQI method (as described in chapter 5; Snijkers, 1997; see appendix 6.1 for the CAQI questionnaire). These interviews were think-alouds, including follow-up probing. In addition to thinking-aloud, with a number of questions meaning-oriented probing was used to investigate the meaning of some concepts in the questions in more detail. Also a test question was added to the end of the questionnaire to elaborate on the key question, i.e. to find out whether no items had been forgotten. This pre-test protocol was chosen because of the research questions. However, this made pre-testing on the phone impossible.

Those in-depth interviews were conducted using the Blaise questionnaire of the Continuous Survey on Living Conditions (in Dutch: *Doorlopend Leefsituatie Onderzoek DLO*). In this questionnaire, the existing questions on income had been replaced by the newly designed set of questions. An interview started with questions on the household composition, followed by questions on housing, time use, employment, other activities, working conditions and looking for work. These questions were asked according to standard procedures of face-to-face field interviews, without thinking-aloud and probing. Then, the questions on education level were asked. With these questions, the respondents were trained to think-aloud, meaning that they were asked to tell all their thoughts while hearing the question or coming to an answer. And after this, the actual in-depth interview was conducted with the new income questions. This completed the interview. (See appendix 6.1 for the CAQI questionnaire.)

The in-depth interviews were divided into two parts, immediately following each other. In the first part, the income questions were posed to the respondent one by one (according to the correct routing) using thinking-aloud. Also in this part, meaning-oriented probes were asked with the following concepts:

- ‘the person in the household with the highest net income’ (in question *HdBreadW*, see appendix 6.1),
- ‘an ordinary month’ (in *IncMonth*),
- ‘the last 12 months’ (in *IncYear* and *IncYearS*),
- ‘about this year’ (in *TaxRturn*).

In the second part, the key question (the total net household income) was investigated in more detail using a test question. Here, the respondent was asked to fill in a scheme, in order to calculate the household income, while the original question was repeated. With this test question we tried to find out whether the question had been fully answered in the first place, i.e. to find out whether some aspects with regard to income had been forgotten. A meaning-oriented probe on ‘net income’ followed this fill-in scheme.

Since the key question was worded in three different ways to cover three groups of Dutch populations, volunteering respondents had to be recruited in such a way that those three groups were represented. All together, 31 volunteering respondents were selected with the following characteristics:

- self-employed (10),
- employed, with a monthly income or yearly income (21).

In advance, they were informed about the special character of the interviews.

Ten shopkeepers (self-employed) were selected. With respect to this group, there was only one selection criterion: receiving an income from their own business. The in-depth interviews with this group of respondents were conducted at their homes or in their businesses.

Those employed (all together 21), were selected on the basis of three background criteria: gender, age and education level. The distribution of the respondents with regard to these background characteristics is shown in table 6.1. The interviews with these respondents were conducted at the office of Statistics Netherlands in Heerlen by researchers from the cognitive laboratory[1]. The question on the total net *monthly* income (here after called the *month* question) was asked to 16 respondents. The other 5 respondents got the question on the total net *yearly* income (the *year* question), meaning that in these interviews the correct routing was not followed.

After the interviews had been conducted, the distribution of respondents for these variables was a little different from what was expected beforehand (see table 6.2). Two respondents, who were self-employed, were not the head breadwinners. Their spouses received ‘wages or salary’, which meant that the month question was asked to these two respondents. As for those respondents who were asked

[1] Martin Luppes and the author.

Table 6.1. Number of respondents for being (self-)employed, age, gender and education level

Age	employed						self-employed	
	education level male			education level female			male	female
	low	middle	high	low	middle	high		
18 -< 25	-	1	-	1	-	1	-	-
25 -< 45	2	3	3	2	1	2	4	1
≥ 45	2	-	-	1	1	1	4	1
Total	4	4	3	4	2	4	8	2

Table 6.2a. Number of respondents, the month question was asked, with regard to: being (self-)employed, age, gender and education level

Age	employed						self-employed	
	education level male			education level female			male	female
	low	middle	high	low	middle	high		
18 -< 25	-	1	-	1	-	1	-	-
25 -< 45	2	1	3	1	1	1	-	1
≥ 45	1	-	-	1 *	1	1	-	1
Total	3	2	3	3	2	3	-	2

* both month and year question were asked to respondent; accent on month question.

Table 6.2b. Number of respondents, the year question was asked, with regard to: being (self-) employed, age, gender and education level.

Age	employed						self-employed	
	education level male			education level female			male	female
	low	middle	high	low	middle	high		
18 -< 25	-	-	-	-	-	-	-	-
25 -< 45	-	2	-	1	-	1	4	-
≥ 45	1	-	-	-	-	-	4	-
Total	1	2	-	1	-	1	8	-

to answer the month question, one respondent did not know the monthly income. Therefore, the year question was asked to this person. However, although the year question was asked, during the interview, the month question was investigated in more detail with regard to interpretation. (In table 6.2 this person is marked with '*'.) So in the end, the month question was investigated with 18 respondents, the year question with 5 respondents, and the year question for self-employed was asked to 8 respondents.

All interviews had been audiotape recorded. The interviews had been analysed by the researchers using the form in figure 4.1 of chapter 4. Each researcher coded all interviews. Then, the results for each question were combined into one concluding form.

6.3. Conclusions and suggestions

In the Netherlands, income is considered a sensitive topic. Despite this fact, almost all respondents co-operated during the interviews and reported their income. This may have to do with the special character of the interview. The in-depth interviews resulted in a lot of information with respect to the question-and-answer process. First of all, the pre-test made clear that in the questions five key concepts may be distinguished. These concepts are:

- 'the head breadwinner',
- 'the major income source of the head breadwinner',
- 'the net income for this household', i.e. the total net household income,
- 'the net income' in itself, and
- the period for which the income has to be reported: 'monthly' or 'yearly'.

The first two concepts refer to the first two questions in the set. These questions did not give rise to any major problem, although the question on the major income source of the head breadwinner is quite long, due to an extensive list of response alternatives.

The others three concepts have to do with the key question on net income. The pre-test resulted in a number of difficulties with regard to this question. The major difficulties are:

- reference set change, complicating a correct interpretation of the key question;
- proxy reporting of income, making retrieval of the necessary information difficult, and thus resulting in best guesses instead of accurate and reliable answers;
- hard to answer with regard to retrieval of the needed information, especially for self-employed; and
- ambiguous interpretation of the three key concepts in this question, resulting in invalid answers.

The first problem in the key questions has to do with a change in reference. After the introduction was read aloud (in which was put forward that the following questions deal with the income situation of *the household* the respondent belongs to), first the two questions on *'the head breadwinner'* and *'the major income source of the head breadwinner'* were asked. Then, the key questions on *'the total net household income'* was put forward, instead of asking for the actual income of the head breadwinner, which logically, would have been the next question. (See appendix 6.1 for the question wording.) Thus, in the introduction the reference frame is set to the household. In the following two questions, the reference frame is restricted to the head breadwinner and his/her major income source. The next question, however, deals with the household again, suddenly widening the reference frame. Because of this sudden change in perspective, incorrect answers are more likely to occur. The change in perspective is in contrast with the reference frame the respondent is building up during the interview, and thus may cause errors in the interpretation

of the question (see Clark & Schober, 1992, p. 40[2]). In field interviews the change in perspective (from head breadwinner to household) may lead to underreporting, since it is not clear when these errors occur.

The in-depth interviews showed that in some cases, respondents reported only the major net income of the head breadwinner, instead of the total net household income (and e.g. not including their own income), even though the instruction was to sum up all earnings. The question ends with an instruction (immediately after the actual question sentence), saying that all incomes had to be added together. It seems however, that respondents were already trying to find an answer to this question in their mind, without listening carefully to the instruction. Besides, we may infer that the length of the question seemed to hinder a correct interpretation too. Once, the change in reference was made clear to them (by repeating the question or with help of the fill-in scheme), they tried to answer this question for the 'household'.

Another problem that appeared with this question is proxy reporting. In the interviews, it appeared that reporting the income of the head breadwinner by other household members is difficult and not very reliable. Especially children in the household and spouses showed difficulties in retrieving the income of the head breadwinner. And in addition, we may conclude that proxy reporting of income appeared to be difficult in general, e.g. the head breadwinner having to report the small income of his children. This may have to do with the fact that in the Netherlands income is considered a sensitive topic and is not discussed openly, even within a household between spouses.

Apart from the change in reference and proxy reporting, the question on the total net household income was not always easy to answer, with regard to retrieving the needed information. Especially self-employed respondents found it difficult to come up with an answer right away. They said that they had to check their administration, annual account or contact their accountant. Employed respondents came up with an answer fairly easy. They indicated that they visualised their payment slip, income tax form, bank or postal account slip.

With regard to interpreting this question, some problems came about. One of these problems had to do with the interpretation of '*the net income for this household*'. These words are supposed to indicate that all net incomes of all household members (in a particular period of time) have to be added together (as is also made clear in the instruction), resulting in the total net household income. Some respondents however, interpreted this concept as the income that was used for

[2] Clark and Schober state (p. 40): "*People interpret successive questions as related in topic - unless they are told otherwise.* Ordinarily, survey questions either continue the current topic of conversation or introduce a new topic. When a question is heard as continuing the current topic, it should be interpreted as related to the previous question, and it is."

paying all household expenditures (see Clark & Schober, 1992, p. 28[3]). This interpretation was especially the case for odd household compositions, like not-married people living together, small communities, living-in relatives (brother, parents). But also in the case of children who have an income of their own and who are still living with their parents. Thus, resulting in underreporting, since incomes were not included or simply forgotten. As for single respondents or those who were married (or had a steady relationship), with any children who did not have an income of their own, this concept did not give rise to diverging interpretations, since both income concepts are one and the same.

The concept '*net income*' in itself was clear to employed respondents. They interpreted it unambiguously and in accordance with the view of the questionnaire designer. Almost all of them reacted spontaneously by saying: "the money I get in hand" or "the money I get on my bank (or postal) account." Those who were self-employed however, interpreted the concept 'net income' in different ways. They referred to net income as: "profit after subtraction of all taxes", "profit after subtraction of all taxes and other financial obligations (like insurance premium, creditors, electricity)", or "monthly private transfers from the business." To self-employed respondents it was difficult to describe this concept. And furthermore, they used the same words for different concepts, which makes it very difficult to get one unambiguous interpretation of 'net income'.

As for the period for which the income had to be reported (monthly or yearly), it appeared that respondents did not always report the income according to the period mentioned in the question. Some respondents came up with the income according to the period of payment, regardless of the period mentioned in the question. E.g., some respondents reported their 4-weekly income, despite the fact that in the question was asked for the monthly income. Some of them only mentioned this income, while others turned this into the monthly income. Also for the question on the yearly income, this problem appeared: some respondents (both employed and self-employed) spontaneously mentioned their monthly income. Thus, respondents tend to report the income for the period for which they are paid. With regard to the information-processing task (retrieval and judgement), this indicates that they use the information that is easily available to them, based on their own reference frame. They try to make to make it easy on themselves, without making the requested calculations (see Clark & Schober, 1992, p. 35[4]).

Furthermore, the mere fact that in the question a period is mentioned, sets the reference frame for respondents with regard to the income sources that are to be retrieved. In this way, respondents are being focussed on earnings in that particular period, while forgetting about other sources. E.g.,

---

[3] Clark and Schober state (p. 28): "*Respondents fail to see when the surveyer is using words differently from the way they use them.*"

[4] Clark and Schober state (p. 35): "*Respondents estimate factual answers that would take too long to figure out precisely.* Respondents tacitly reason, 'The surveyer is asking me a question I should be able to answer immediately. So even though he seems to be asking a precise answer, he couldn't be because he couldn't possibly expect me to compute it in the time available. So I will make a best guess.' This way precise questions get less precise answers than surveyers expect."

payments on a yearly basis, like savings-interest revenues or vacation benefit were forgotten with the month question. Thus, both for self- and proxy reports, retrieval of all information seemed difficult. Again, this may cause underreporting. And in case these earnings have not been forgotten, this complicates the judgement process, since respondents have to calculate the monthly equivalent of these yearly earnings.

In the question on the yearly income, the following wording is used: '*the net income in the last 12 months*'. Reactions from respondents showed that *'the last 12 months'* is interpreted differently than when this period is linked to income. 'The last 12 months' is interpreted as "from now going back 12 months" or "the period March 1993 - March 1994". In connection with income however, this period is interpreted as "the last calendar year: 1993" or "the months January - December". They remarked that it was easier for them to come up with the income in 1993 and that there are no big differences between the income in 1993 and the income in the period March 1993 - March 1994. Calculating the amount as requested would be too difficult to do (see Clark & Schober, 1992, p. 35[5]).

The concept '*the income in an ordinary month*' did not give rise to any major interpretational problems. With regard to this concept, respondents talked about the income they received after the usual hours of labour (in cases in which the income was dependent on the number of working hours), or in case of a fixed monthly income, this income was mentioned, without any additional payments.

The key question was followed by a question on health insurance: whether health-insurance premium had to be paid still from this income. This question resulted in interpretational problems for those respondents who had to pay premium for one person in the household and not for others. This indicates that this is a double-barrelled question. Furthermore, it appeared from the interviews that respondents with the same kind of insurance reported different answers, indicating that this self-report may not be very reliable.

At the end of the interview, respondents were asked whether they expected to receive a tax return for this year. The intention of this question was to find out whether they expected to get money back in 1995 based on their income tax of 1994. Almost all respondents however, based their answer on the tax year 1993: the year for which in 1994 the fax form had to be filled in. While probing into the meaning of '*in this year*', respondents replied: "This year?! This year is 1994". However, while answering the question they were thinking about the tax year 1993. So, the question was interpreted as: "Do you expect to get money back from your income taxes this year?" or "Do you expect to get money back in 1994, based on 1993?" When subsequently asked how much they expected to get back, they indicated that it was very difficult to come up with a

[5] See footnote 3.

particular amount of money, although they had only recently filled out the tax form for 1993. In case the question was interpreted correctly, it was absolutely impossible to answer this question.

In summary, 'the total net household income' was not always what it was meant to be by the researcher, namely the total of all earnings received on the bank account, for all sources and or all persons who share their daily life. For self-employed and for odd household compositions in particular, one may expect diverging reactions. As for self-employed respondent, the questions are hard to answer without consulting their administration or accountant. Employed respondents may very well be able to answer the questions on top of their heads for themselves. But, keeping the change in reference and proxy reporting in mind, with all effects combined, the answers to this set of questions may not be accurate, valid and reliable. In face-to-face field interviews, this may lead to underreporting of income. In telephone interviews, the effects may even be stronger, since the time pressure to answer is greater than in face-to-face interviews. As a consequence, some households will be assigned to lower income categories.

These effects, as found in the face-to-face in-depth interviews, are quite obvious with respect to the question-and-answer process. Therefore, we suggest the following changes:

- As far as the change in reference is concerned, it is recommended that the questions are reworded completely, and thus avoiding changes in perspective. Leave the questions on the head breadwinner out (since they only serve as a routing questions) and ask directly for the net income. At the same, time try to make the key question shorter. The question is quite long as it is, especially for telephone interviewing.
  Another solution is trying to make the changes in perspective more evident to the respondent. However, this would not really solve the problem, and would very likely make the key question even longer.
- The concept 'household income' should not be used. This concept should be described as e.g. 'all incomes from everyone living here in this house'.
- Asking for proxy information is a very hard job for respondents and the information is not very reliable. Thus, getting the income from other people in the household by interviewing only one person should be avoided as much as possible. However, this would complicate the face-to-face and telephone interview. Therefore, we suggest making it a closed question, with not too detailed income categories as response alternatives. This would also come close to one of starting-points with regard to the key question, as is also pointed out in the introduction to the respondent. The question asks the respondent to be precise in his answer, while the answers are only used for clustering households. A drawback of predefined answering categories however, is that they are fixed and afterwards no other clustering may be chosen.
- It is recommended to ask the income for the period in which the respondent receives it, and not with predefined monthly or yearly periods. Afterwards the income can be recalculated to meet a standard period.
- The concept '*net income*' can be described using the words of the respondents, e.g. as "the earnings everybody receives on his/her bank or postal accounts".

- Leave the questions on health insurance premium and tax returns out. However, the question on tax returns may also be reworded: “Do you expect to receive any tax returns next year on this year’s income?” without asking for the exact amount.

It is up to the researcher to see whether these changes are in accordance with his views on how to measure this concept. However, with respect to respondent friendliness, we feel that these changes would make the questions easier to answer for the respondent and thus resulting in valid data on income. As for newly worded questions, it is recommended that they are pre-tested again. Only then, it may be concluded that the changes actually are improvements.

6.4. Results of the cognitive interviews

In this section the results of the 31 cognitive interviews on which the conclusions are based, will be described. On request of the Department of Social Accounts, respondent reactions will be reported in full detail for each question separately.

6.4.1. Head breadwinner (HdBreadW; see appendix 6.1 for question wording)

For most respondents, it was not very hard to come up with the person in the household who had the highest net income. The respondents knew who the head breadwinner was, even if the respondents was not the head breadwinner him/herself. (It should be noticed that the question does not ask for the name of the head breadwinner, but only asks for indicating whether the respondent is the head breadwinner or not.) In table 6.3 the results of this questions are presented.

Table 6.3. Head breadwinner for income situation

Head breadwinner in the household	Income situation			Total
	Monthly	Yearly	Self-employed	
Respondent him/herself	7	3	6	16
Spouse/partner of respondent	8	2	1	12
One of the parents of respondent	2	-	-	2
Brother of respondent	-	-	1	1
Brother of spouse/partner of respondent	1	-	-	1
Total	18	5	8	31

With 5 respondents, an interesting reaction was noticed while answering this question. Three of them were self-employed respondents; the other 2 received a monthly income. For 2 of the 3 self-employed respondents it was not clear who had the highest income in the household, since both spouses worked in their own business. In the first business, the head breadwinner was determined at the end of each financial year (for tax reasons); in the other business it was absolutely impossible to determine the contribution of each partner.

The household composition in the third business was quite complicated. It was a household of three brothers en their mother. Two brothers were partners in business (one of them was the respondent); the third brother had a business of his own. The respondent replied to this question in the following way: “Maybe, I am, together with my brother, who is my business associate, … Probably, we’ll have the highest income, but that income is not used for … Most household expenditures are paid by my other brother. We pay him rent and boarding. Our income situation has little to do with housekeeping.” As an answer to the question the respondent said that he is not the head breadwinner. From this reaction it became clear that in the answering process the respondent was thinking about the person in the household who is responsible for the housekeeping and the income that is used for paying all expenditures. The answer to the next question therefore was the third brother: he made profit from his business.

As mentioned above, the determination of the head breadwinner for those who were asked the month question, resulted for 2 respondents in an interesting reaction. The first respondent lived in a household composed of himself, his girl friend and her in-living brother. The brother had the highest income, but he was not responsible for housekeeping. Every month he paid 500 Dutch guilders for boarding. In the words of the respondent: “The head breadwinner is the brother of my girl friend. But his salary is not used for running the place: he pays us a certain amount every month. So, he pays us a board of 500 guilders every month, but he earns the highest income.” From other remarks it became clear that he considered his income and his girl friend’s as one. This is what he called the household income, with those 500 guilders added. While probing into their incomes, the respondent remarked that the earnings of his girl friend are higher than his income. From this reaction it became clear that to the respondent the head breadwinner the person in the household with the highest income was, but this person did not belong to the core of the household.

In the last case the household consisted of 4 persons: father, mother and two sons, all having an income of their own. One of the children was the respondent. While answering the question, he remarked: “I think it is my father.” He was not sure, however. In this case, we got proxy information from a respondent who did not belong to the core of the household.

6.4.2. Income source (InSource)

Like with the first question, determining the major income source for the head breadwinner was not very troublesome to most respondents. Still, with 10 respondents an interesting reaction came forward: 3 self-employed, 6 respondents who answered the month question and 1 who answered the year question.

As for the self-employed respondents, the reactions had to do with a change in perspective, the interpretation of ‘profit from own business’ and the extensive response items. With regard to the changes in perspective, the next reaction (from a respondent who is head breadwinner) is

illustrative: “My income comes out of ‘profit from own business’. In this question you are asking for my income only and not for my wife’s, isn’t it?” Here we see that the respondent asked for clarification. Since in the introduction and in the question on the head breadwinner was talked about the household, it was not clear to him whose income was meant. The change in perspective (from household level to head breadwinner) was confusing to the respondent.

As far as the interpretation of ‘profit from own business’ is concerned, one respondent remarked: “Profit out of business, really. It is always afterwards. You get earnings every month; you live now and here, but actually your income is always determined afterwards, what your income was, after a year, after all taxes and so on have been paid. But, you live from your business, really. You can only live from the profit out of your business, can’t you? This respondent was thinking aloud, trying to find out what his income source really was.

As for the third self-employed respondent, reading all response items was not necessary. While reading the response items he interrupted (with ‘from pension scheme, old age benefit, widow benefit’), saying: “‘profit from own business’, now you do not have to read all items. I do not have other income sources.” Here we may add, as from our experience as an interviewer, that this question was hard to read aloud, due to the extensive response alternatives.

The reactions from the 6 respondents who had to answer the month question had to do with: having two income sources (2), the wording of the question (2), and providing proxy information (2).

Immediately after hearing the question, 2 respondents remarked that they had two income sources. One of them got wages and half-pay as income. He did not know what income was the highest. The other respondent had a business of his own, but was also working in paid employment for another business. He was quite sure that the wages he received was his major income source. The first respondent showed some difficulties in retrieving the necessary information from his memory. The second respondent had no problems with this.

With regard to the wording of the question, one respondent remarked the following: “I think it is ridiculous that even nowadays the word breadwinner is used. Actually, everybody should be able to take care for him/herself, without this being related to the person you live with.” Another respondent was irritated by the fact that ‘social benefit’, ‘half-pay’ and ‘scholarship’ were put together into one response alternative. She said: “Why are these put together? Why is early retirement put apart?” She received a social welfare benefit and didn’t feel any affinity to people who were having half-pay or scholarship.

The respondent who didn’t happen to know his monthly income (and therefore had to answer the year question), could retrieve the correct information in order to answer the question on income source, but he did not know what category to choose. He remarked: “My income is from former

labour. I get two benefits because of occupational disability. Because of these two benefits, I am head breadwinner." Then the interviewer repeated the response items one by one. The respondent didn't know what to choose. In the end he chose the rest category 'another income source' (despite the fact that his benefit is a 'social benefit').

As for the last 2 respondents to be described here, they did not belong to the core of the household. These proxy respondents knew the major income source of the head breadwinner, but they added that they didn't know the exact amount.

6.4.3. Does not know the net household income (NoNetInc)

The 2 respondents, who did not belong to the core of household, were asked whether they know the net household income approximately. One of them answered this question, thinking of the head breadwinner only: He didn't know exactly what his father earned. He guessed and mentioned a certain number. Due to the change in perspective (from head breadwinner to household level), this respondent was put on the wrong track.

The other respondent could not come up with an answer at all: she really didn't know. From this we may conclude that it may be very hard for respondents to provide proxy information. (Actually, here, this interview would have ended. To get some information from this interview, the next questions were answered for the respondent herself.)

6.4.4. Income

In this subsection the respondent reactions with regard to the income question will be described. First, the question on the yearly income for self-employed will be discussed (in subsection 6.4.4.1), followed by the question on the monthly income (subsection 6.4.4.2) and yearly income (subsection 6.4.4.3) for those working in paid employment. In every subsection, the following items will be discussed: general answerability of the question, respondent reactions, the financial sources that were mentioned to get to know the necessary financial information, the interpretation of 'net income', the interpretation of 'the last 12 months' or 'an ordinary month', and reactions with regard to the test question (the fill-in scheme in the second part of the interview)[6].

[6] In the in-depth interviews we also asked the respondents to comment on the accuracy of their answer and the sensitivity of income. These probes however, did not result in usable data. As for the accuracy, respondents reported their best guess: when they would have thought otherwise, they would have said so. And as for sensitivity, this is a hypothetical question, and is not very predictive as to the behaviour of respondents in the field.

6.4.4.1. Yearly income for self-employed (YearIncS; self-employed)

Not all self-employed respondents were able to *answer the question* on the net yearly income in the last 12 months easily. This is shown in table 6.4. Three out of the 8 respondents came up with a straightforward right away. (There was only one income in these 3 households: 2 respondents were single and the spouse of the third respondent did have an income of her own.) The other 5 respondents said that it was hard for them to come up with an amount for themselves or for their spouse. Some of the reactions were:

- "That is a good question! I really don't know what my wife makes a year. I have to estimate the amount."
- "For me, as a starting entrepreneur, it is hard to get a picture of that."
- "Each month we take an amount from our business as our monthly private earnings. But I don't know what I have made in total. And certainly not in the last 12 months: we haven't yet closed the financial year 1993, we did so for 1992. But I don't have those numbers in my head."
- "I cannot say: I have to look for it my annual account. I don't know whether I am supposed to? It is hard to say, since it changes every year. I don't mind looking it up for you, that is no problem at all."
- "Well, that is hard to say. I think, I should get my papers. I'd say, just leave it open, I can't say."
- "It is very hard to say when you have a business of your own. (…) I have to figure that out. I don't know that by heart. I have to get my administration."

For this last reason, two respondents did not come up with a figure at all.

Table 6.4. Answerability of question on yearly income by self-employed

Whose income is this:	Respondent knows income:			Total	Respondent has no spouse	Spouse has no income
	Yes	Approximately / hard to say	No			
Respondent	3	3	2	8		
Spouse	-	2	1	3	3	2
Other housemate	-	1	-	1		

Four respondents asked for clarification. This was about:

- whether the respondent was supposed to look for data in the administration (1 respondent);
- whether the respondent was supposed to subtract income taxes from the income (1 respondent);
- whether the question is about the income on a yearly basis (2 respondents).

With regard to this last aspect another respondent (who each month got private transfers from his business) was convinced that the question was about the monthly amount, instead of the yearly income. He came up with his answer immediately after the question was read aloud. After probing into the yearly amount he said: "But that is not what is asked for, is it? You asked for the monthly amount." Then the question was read aloud again, and he gave a yearly amount, being the monthly private transfers multiplied by 12. So, his answer was based on the information available to him at that time, and not the information needed to give a valid answer to the question.

While answering the question, respondents spontaneously mentioned several *sources* that could be consulted to get the exact amount. These sources are listed in table 6.5.

Table 6.5. Sources mentioned for financial data

Source for financial data	Number of times mentioned *)
annual account/annual returns/annual report	6
book keeper/accountant	3
tax form	2
administration/book keeping	2
spouse / partner	1
no source specified: own memory	1

*) There are duplicates.

In the key question the concept '*net income*' is used. As from the respondent reactions we may conclude that it was difficult for self-employed respondents to describe this concept. And furthermore, they used the same words for different concepts, which makes it very difficult to get one unambiguous interpretation of 'net income'. Here we have listed some reactions of respondents after probing into the meaning of 'net income':

- "That is the profit from my business at the end of the year. The amount I find on my annual returns. In fact, that is my net income. Then I am free to go. Then I have paid all my debts, including health insurance premium and so on. The amount that is left over at the end of the year is my net income."
- "The amount that is left over after paying income taxes, so the money that can be spent freely for businesses, mortgage, housekeeping, hobbies, etc. A financial obligation to manufacturers, taxes and the bank is money I cannot spent freely. This is already settled before I get the income that can be spent freely."
- "The net income is the profit from my own business after all charges with regard to the business have been deducted. Then you get a certain amount, from which social charges and so on have to be subtracted: social charges and tax. This is also deducted and then you get the net income."
- "The net income is the net profit from my business minus income tax, property tax and social charges."
- "The money that is left over after everything has been paid: taxes, electricity, creditors, etc. So, all goods you have bought. That is why it is so difficult. Because, there are always items that are still open at the end of the year. The net amount is the amount that is left over after I paid everyone. Then I see what is left over. That is mine."
- "The net amount is what is clearly left over, when I have paid all my obligations: insurances, personal obligations, and income tax. The amount I have to life from. That is 'net' to me. Then all encumbrances like all personal insurances (like health insurance), taxes, rent, gas, water, light have been paid. So, what is really left over."
- "My net income is the private amount we get every month from the business. That is the amount we live from. The rest is dealt with in a businesslike manner."

With regard to the interpretation of the concept '*the last 12 months*' there was a difference with 'the income in the last 12 month'. In general, 'the last 12 months' was interpreted as "from now going back 12 months", "from now to March last year". When related to income the interpretation became "the financial year", "January up to and including December", as is shown in table 6.6. Some reactions were:

- "'The last 12 months'? Well, actually I have counted from January until December. Why, is that a problem? Look, on average the amount is the same, whether I count from March 'till March or from January 'till January. But, for me it is easier to count this way. On the 31st of December I draw up the final balance sheet."
- "For us, 'the last 12 months' is always the financial year: January up to and including December. It is hard for me to say what I make in a month, because it is hard to determine what the profit from the business is each month.
- "'The last 12 months' is from now going back 12 months. But, that is impracticable for me to do. I guess, this question is hard to answer for self-employed. I think that every business is somewhat behind."

Table 6.6. Interpretation of 'the last 12 months'

Interpretation	'the last 12 months'	'the income in the last 12 months'
March - March (now going back 12 months)	5	-
January - December (financial year)	2	4
1993	1	1
1992 (as estimate for 1993)	-	3

As for recalling the income of other members of the household (proxy reporting) and the interpretation of 'the household income', 4 self-employed respondents provided important information with the *fill-in scheme*.

We have mentioned before that it maybe hard for the respondent to come up with proxy information. Once again this became clear from the reaction of the respondent who lived together with his two brothers and his mother. He said: "Then I am supposed to add up the net incomes of the four of us. I do not know the income of my third brother very precisely. I do know the old age benefit of mother approximately. I do not know whether I could estimate the income of my third bother. I guess that would not be very realistic. That is none of my business." In the end he filled in his net income, the net income of his second brother (his business-partner), and the income of his mother. He didn't fill in the income of his third brother, so that the total amount cannot be calculated. (When this respondent answered the question on the total yearly net income for the first time, he came up with an estimate of his income and that of his brother in business.)

With help of the fill-in scheme, it appeared that with regard to '*the total household income*' two self-employed respondents had forgotten to count the income of their living-in child. They remarked:

- "I am the head breadwinner. There is only one income in this house, which is mine. My daughter has an income of her own. That is none of my business. She even doesn't pay any board. She is still living with us, but she has to take care of herself."
- "As I have said before, to see what mine and my wife's incomes are I would need to consult my administration. We both have half of the income from our business. And furthermore, … my son has nothing to do with our business. He has a business of his own and is taking care of himself. The only thing is that he is still living with us."

The respondents did not count the incomes of the living-in children as being a part of the household income.

With help of the fill-in scheme, another self-employed respondent recalled that apart from his income from his business he also got deposit revenues. He said: "Actually, this is not what you think of, when talking about income. It is not on top of your mind. When someone would ask me to report my income, I would not be thinking: 'Oh yeah, and I also have deposits'. You see it when you get your bank slip. It is silent money; it is put away. When someone would ask me how much money I make, then I would think about profit or wages in the first place."

6.4.4.2. Monthly income (MnthInc)

Seventeen out of the 18 respondents, the question for the monthly income was asked, were willing to report their income. For a number of them, *retrieving the correct information* was not an easy task to accomplish, as is shown in table 6.7. The one respondent, who refused to give her income, was not willing to say anything about it. She indicated that she knew her income and that of her partner. This indicates that to some respondents income is a sensitive topic that is not discussed in public. Therefore, we leave this respondent out of our analyses, meaning that the results that will be described below are based on 17 respondents.

Table 6.7. Answerability of question on monthly income

Whose income is this:	Respondent knows income:			Total	Respondent has no spouse	Respondent/ spouse has no income
	Yes	Approximately / hard to say	No			
Respondent	12	4	-	16		1
Spouse	7	3	1	11	4	2
Other housemate	1	2	1	4		

The respondents, who indicated that this question was hard to answer or did not know the income, mentioned the following reasons:

- the income changes every month (this was the case for 1 respondent and for 2 partners);
- the income is paid in 4-weekly periods and therefore has to be turned into a monthly income (this was hard for 1 respondent);
- the half-pay procedure is not clear: the respondent did not know how much the amount should be and whether the correct amount had been paid (1 respondent told this about his own income and estimated his income; another respondent didn't know the income of her husband);

- respondent is child of the head breadwinner and therefore did not know the income (1 respondent estimates; another did not have the faintest idea);
- income is very private (1 respondent clearly found it very difficult to mention the income: she thought of all sorts of reasons not to mention her income (from her shop), her husband's income (head breadwinner, on pension) and her son's income (who is helping out in her shop)).

Fifteen respondents indicated that they had to make some calculations in order to come up with the correct answer. For 5 out of these 15, it was simply adding up the individual amounts, which did not give rise to any problems. For the other 10, the calculations were more troublesome, because:

- the income had to be converted from a 4-weekly or 1-weekly income to a monthly income (3 respondents);
- it was not clear what had to be counted (2);
- respondent could not recall the income (2);
- the income varied every month (1),
- there are a lot of numbers that had to be added up (1);
- it was not clear what the income actually is because of the continued absence of information from the paying organisation; the respondent had to estimate (1).

Four out of these 10 respondents asked for clarification:

- "Do I have to include travelling allowance?" (The interviewer replied with: "This question is about the net income as you see it." Then the respondent said: "In that case, travelling allowance has to be included, since it is part of my income.")
- "Oh, I have to include the incomes of the others as well?" (At first, this respondent had only mentioned the income of the head breadwinner.)
- "I don't know whether my scholarship has to be counted as income as well?" (This respondent counts his scholarship in, without waiting for reply. The reaction is quite remarkable since in the preceding question (on the major income source of the head breadwinner) scholarship is mentioned as one of the possible income sources.)
- Another respondent requests for repetition of the question.

The periods, in which the incomes are being paid, are shown in table 6.8. In this table a number of double counts have been processed (indicated by "*"). Three respondents had two incomes: 1 respondent got both incomes monthly, the other 2 got one income each month and the other income on a 4-weekly basis. There was one spouse with two incomes: a monthly and a 4-weekly income; another spouse got his income paid in parts: a weekly fixed advance and a 4-weekly additional income (overtime). One respondent added all monthly incomes of himself and his housemates (father, mother and brother) up.

Eight out of the 10 non-monthly incomes have been converted by the respondents: 6 4-weekly incomes into monthly incomes, 1 1-weekly income into a monthly income and 1 1-weekly into a 4-weekly income. In the end, 3 4-weekly incomes remained as they were.

Table 6.8. Number of incomes for period of payment

Whose income is this:	Period of payment				Double counts (indicated by *)	Total (no doubles)
	monthly	4-weekly	weekly	unknown		
Respondent	14 *	4 *	1 *	-	3	16
Spouse	8 *	4 *	1	-	2	11
Other housemate	5 *	-	-	1	2	4
Converted into:						
4-weekly income			1			
monthly		6	1			

While answering this question, the respondents mentioned several *sources* where to find the necessary financial information. These sources are listed in table 6.9.

Table 6.9. Sources mentioned for financial data

Source for financial data	Number of times mentioned *)
no source specified: own memory	7
payment slip / wage slip / pay-roll	7
tax form	3
bank account slip / postal account slip	2
labour-contract	1
spouse / partner	1

*) There are duplicates.

In contrast to respondents who were self-employed, the concept '*net income*' was unambiguous to people who were in paid employment. 'Net income' was interpreted by 12 out of 17 respondents as: "What is paid into my bank account every month" or "The money I get in hand." Eleven respondents said: "The gross income after taxes and social charges have been paid." Out of these 11, 6 respondents replied in the second instance (spontaneously or after probing) 'the bank account' or 'in hand'. One respondent said it in the following way:

- "Just your wage after social charges and tax have been subtracted. So in fact the money I get on my bank account every month. That is my net income."

Or, as remarked by yet another respondent:

- "Everything that is left from your salary after all deductions have been paid. This is the income you will find on your bank slip. You won't see your gross income."

One respondent even mentioned three different interpretations of 'net income'.

- "The net income is the income after all taxes and social charges have been subtracted. That is what you will find on your pay-slip. This is also the amount that is paid into your bank account."

These interpretations have been summarised in table 6.10.

Table 6.10. Interpretation of 'net income'

Interpretation	Number of times mentioned *)	
What is paid into my bank account. What I get on the bank. The money I get on my bank account each month.	9	12
The money I get in hand.	5	
gross income minus taxes and social charges (health insurance premium, unemployment insurance premium, pension scheme premium)	11	
pay-slip	3	

*) There are duplicates.

The question deals with income in '*an ordinary month*'. Also this concept is investigated in more detail. From the respondent reactions we may conclude that 'ordinary' is interpreted as 'what is normal' to the respondents with regard to the income in a month or hours of labour. 'Ordinary' gets the meaning of 'normal', 'usual', 'on average', 'customary', 'not exceptional'. When talking about 'an ordinary month', some respondents think about the period in the calendar of "4½ weeks" or "30 to 31 days". All interpretations are listed in table 6.11.

Table 6.11. Interpretation of 'an ordinary month'

Interpretation	Number of times mentioned
'usual' hours of labour in a period of a month / 4 weeks	6
'usual' income in a period of a month / 4 weeks	5
'usual' period in the calendar of 4½ weeks / 30-31 days	4
no plain interpretation mentioned	2

Some respondents remarked, while talking about the 'usual' hours of labour:

- "No overtime, haven't worked any additional days, and so there are no extra earnings."
- "When my husband is not working overtime too much. Then there is not much extra payment."
- "Just working for a whole month without leave or vacation (since I won't get paid for days off work) and no long days. Just working 20 hours a week (according to my labour contract) for the whole month."

From these reactions, it seems that the interpretation of 'an ordinary month' is related to 'the usual' hours of labour in a month without any extra payments. In these case the income is dependent on the hours of labour.

Respondent, who were thinking about the 'usual' income in relation to 'an ordinary month', remarked:

- "I get my salary every month, without money for child care, vacation benefit and temporary earnings. All months are the same. That is not changing anymore: I am at my maximum."
- "An ordinary month is a month without temporary earnings, no vacation benefit, no extra payments. Just an average month, when thinking about my income."

The interpretation of the concept 'ordinary month' therefore is dependent on the way the salary is being determined: on the basis of the hours of labour or a fixed monthly salary. As for the

income this means that for all cases the 'usual' monthly income was reported: the income without additional payments, like overtime payment, vacation benefit, benefits, etc.

After the questions were posed according to the usual routing, the question on the monthly income was repeated using the *fill-in scheme*. As for 13 respondents, this scheme did not result in additional information. For 4 respondents however interesting reactions came about.

It appeared that from one household both members were recruited to participate in this investigation (boy and girl friend). Their reactions once again indicated that proxy information with regard to income is unreliable. While filling in the scheme, the girl friend mentioned her own income and a small (irregular) income of her boy friend. (She had mentioned these incomes also the question was posed for the first time.) Her boy friend mentioned his girl friend's income and his scholarship. And furthermore, he mentioned his small (irregular) income. (He did not mention this the first time.) Thus, we may conclude that the girl friend forgot to mention her boy friend's scholarship, both times the income question was posed.

With regard to the interpretation of '*the total net household income*', the fill-in scheme resulted in additional information with 2 respondents. And furthermore, some additional (small) income sources were mentioned, which had been forgotten in the first place. While filling in the scheme, these two respondents remarked:

- "What is the household income: the income you need to keep the household running. You are asking for the net income of the household. In that case you have to ask yourself: How much money is used for housekeeping? E.g., 6 persons make 2000 guilders each, which makes 12000 guilders all together. But, those 12000 guilders will not be used for housekeeping, but let's say only 6000 guilders. The total income of all people living in a community is the net income of all people living there. However, you better not use the word 'household'. To me, 'household' has to do with all the money spent for household expenditures. And I am living in a small community together with my girl friend and her brother. My income and my girl fiend's income are used for housekeeping, with 500 guilders from her brother added. I do not see the rest of his income; I do not have to take that into account. The rest is no part of the household income.
  And, oh yes, I forgot to mention a small income of mine of about 400 guilders each month."
- "You ask me to tell you what I understand by income: everything I receive, both in kind and in money. E.g., I could have made arrangements for paying my interest revenues each month. I do not count my temporary earnings as well, since I do not see that each month.
  My son has a small income of his own: a studentship, but I'm not sure how much that is, I would have to look it up. When you ask me who has an income of his own, in our household, should I add my son's studentship to my income? If that is the case, then the question was not clear to me in the first place. But I do recall now, that you have mentioned studentship. So, I have forgotten to add my son's studentship.

From these reactions we may conclude that incomes that are not used for housekeeping and for paying household expenditures, have not been included in the first place.

6.4.4.3. Yearly income (YearInc; MnthInc = doesn't know)

As said before, the question on the yearly income was posed to 5 respondents immediately. It appeared that these respondents received their income on a monthly basis. In table 6.12, the information that was gathered during the *answering of this question* is summarised.

Table 6.12. Answerability of question on yearly income

Whose income is this:	Respondent knows income:			Total
	Yes	Approximately / hard to say	No	
Respondent	3	1	1	5
Spouse	2	3	-	5
Other housemate	-	-	-	-

The respondents, who said that it was hard to come up with their income or who did not know, mentioned several reasons for this:

- respondent never had been interested in this topic (2 respondents said this about the income of their spouse);
- respondent had to do calculations (1 respondent about the income of his spouse);
- income changes every now and then (1 respondent about his own income);
- respondent forgot to mention his own income: spouse is head breadwinner (1 respondent).

The 2 respondents, who never did show any interest in the income of their spouse (the head breadwinner), remarked:

- "Each year, oh my! I never went into that. I do not know much about that. I never work that out."
- "I'm not good at that, at those things. What does my husband make? Well, it's quite absurd, but I don't know."

In the end, they both came up with an amount.

As far as calculating the income is concerned, one respondent mentioned that this was troublesome to him. Another respondent calculated the yearly income by multiplying his income by 13 (that is including a month of vacation: vacation benefit). Yet another respondent thought about his tax form, he had filled out recently, and remarked: "When you would have asked me this 6 months later, I wouldn't have known. Then, it would have been necessary to calculate the yearly amount on the basis of my monthly earnings. My income changes every month."

One respondent had heard a completely different question, namely: "What do you spent net each month?" He remarked: "Well, that is hard to say. Because, what is left over net at the end of each month, fluctuates every month. It depends on your spending habits. Well, what is left over on average? It is hard to say." Then the question was repeated, and he said: "That is a long question. It is about the net income in the last 12 months. So, that is the yearly net income." Then he asked

himself: “But how did I get to the spending habits, just now?” Why this mistake came about, remained unclear (even after probing).

While answering this question, the respondents mentioned a number of *sources* that could be consulted to get the necessary financial information. These sources are listed in table 6.13.

Table 6.13. Sources mentioned for financial data

Source for financial data	Number of times mentioned *)
payment slip / wage slip / pay-roll	3
bank account slip / postal account slip	2
yearly report of paying authority	1
tax form	1
spouse / partner	1
no source specified: own memory	1

*) There are duplicates.

And again, we probed into the meaning of ‘*net income*’. Like with the month question (subsection 6.4.4.2), here again ‘net income’ was interpreted by most respondents as “the money paid into your bank or postal account”. The various interpretations are listed in table 6.14.

Table 6.14. Interpretation of ‘net income’

Interpretation	Number of times mentioned *)	
What I get on the bank by my employer. What is paid into my postal account. The money I get on my bank.	4	
The money I get in hand.	1	4
The money I can spend freely.	1	
gross income minus taxes and social charges (health insurance premium, unemployment insurance premium, pension scheme premium)	3	
pay-slip	2	

*) There are duplicates.

Table 6.15. Interpretation of ‘the last 12 months’

Interpretation	‘the last 12 months’	‘the income in the last 12 months’
March - March (now going back 12 months)	4	1
January - December (financial year)	-	2
Income in 1993	-	1
No plain interpretation mentioned	1	1

Also was probed into the meaning of ‘*the last 12 months*’. The various interpretations are listed in table 6.15. Just as was the case with respondent who were self-employed, differences in interpretation between ‘the last 12 months’ and ‘the income in the last 12 months’ were noticed. Again, ‘the last 12 months’ was interpreted as “from now going back 12 months”, while the last 12 months related to income was interpreted as the calendar year. One respondent remarked: “The last

12 months? For me is that from March to March. But, to me the easiest way to get my income is to look for January 'till December (the calendar year). By the way, there is no difference in income."

And again, after all questions had been posed, the *fill-in scheme* was presented to the respondents. For 3 respondents, this did not lead to additional information. For the other 2, the scheme resulted in additional information with regard to the change of perspective and the interpretation of the household income:

- "Well, that is what we haven't talked about yet: about my salary. We talked about my husband's income (= the head breadwinner), but my income has to be taken into account as well."
- "Besides my income and that of my partner, there is also my mother's income. I didn't count that. She pays us a small amount for board, but she is doing everything herself. That remains separated."

For the first respondent the change in perspective (from head breadwinner to all members of the household) was important with regard to her answer. While answering the question for the first time she only thought about her husband's income. Now however, when filling in the scheme, it appeared that she also had in income of her own. The second respondent didn't count his mother's income as part of the household income. As far as he was concerned, her income had nothing to do with his and his spouse's income. Although they lived together in one and the same house, according to him, they had separated households.

Finally, it should be mentioned that, while filling in the scheme, 2 respondents explicitly filled in their monthly income, although the question on the yearly income had been repeated. This indicates that for answering this question one uses the information that is available. It is not self-evident that this information is turned into the data as asked for.

6.4.5. Health insurance (HlthIns)

The question on the whether or not health insurance premium still has to be paid, is asked to 21 respondents who are working in paid employment (those from subsection 6.4.4.2 and 6.4.4.3). (This question was not asked to 1 respondent, because of the routing). This question resulted in a number of conflicting reactions:

- it was not clear whom it concerned (what answer should be given when for one person the premium has to be paid, while this is not the case for another), and
- respondents with (as it seemed) the same kind of insurance (LIASS), came up with different answers.

The various reactions are listed in table 6.16.

Table 6.16a. Reactions for those answering 'YES'

Reaction	Number of times mentioned
YES, I have to pay this every month	5
YES, for me; my partner has a compulsory insurance	1
YES, it is subtracted from my salary automatically; it is on my pay slip	1
YES, a part of the premium for the health insurance fund is subtracted from my salary, another part has to be paid by respondent himself	1
YES, I pay LIASS; my partner has a private insurance; both premiums are paid separately	1
YES, a private insurance for my partner and my children, and a small amount for me to the health insurance fund	1
NO … YES, I have a LIASS insurance; I have to pay that every month	2
NO, not for me; but my partner does pay premium, so YES	1

Table 6.16b. Reactions for those answering 'NO'

Reaction	Number of times mentioned
NO, I have a LIASS insurance; I do not need to pay the premium myself (is being subtracted from my salary)	2
NO, it is the net amount; the premium has already been subtracted	2
NO, is paid by employer	2

6.4.6. Taxation

6.4.6.1. Tax return (TxReturn)

The last two questions were on tax returns. Respondents were asked whether they expected to get a tax return on this year. The intention of this question was to find out whether respondents expect to receive a tax return next year (1995) on the basis of this year's income taxation (1994). However, almost everyone answered this question while thinking about the income taxation of 1993. Only 2 respondents referred in their answer to the expected situation in 1995. The reactions to this question are listed in table 6.17.

Table 6.17. Reactions to taxation question

Reaction	Number of times mentioned
YES, I have filled out the tax form (1993) only recently; this year we get a return for last year	9
NO, not for last year (tax form has only recently been filled out)	6
NO	2
NO, not for last year, and I do not expect any returns for this year	1
YES, last year; I don't know for this year, although I expect some returns	1
YES, I think so: my partner gets a return for last year (1993)	1
Doesn't know	1

6.4.6.2. How much tax return (HowMuchR)

The 11 respondents, who answered the last question with ‘yes’, were asked how much money they expected to get back. Only 1 out of these 11 actually indicated how much he expected to get back in 1995 over 1994, although this was hard for him, since the year only had just begun. He said: “Maybe I could give an amount in October, but not now.” The other 10 talked about 1993 and referred to the tax form (about 1993) they had filled out only recently. Also, for these respondents it was hard to come up with an answer: 5 out of 10 respondents said that they did not know anymore what they had calculated, or they did not know because their spouse had filled out the tax form.

References

Clark, H.H., and M.F. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), Questions about Questions, pp. 15-48.)

Snijkers, G., 1994, What is the total net household income? Cognitive interviews on income questions. Design and Results. (In Dutch: 'Wat is het totale netto huishoudinkomen? Vragenlabonderzoek naar de gecoördineerde verkorte inkomensvraag. Opzet en resultaten. Internal, unpublished CBS report. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1995, What is the total net household income? Cognitive interviews on income questions. (Paper presented at the International Conference on Survey Measurement and process Quality, April 1-4, Bristol, UK.)

Snijkers, G., 1997, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. (Bulletin de Methodologie Sociologique, No. 55, pp. 93-107.)

Tanur, J.M. (ed.), 1992, Questions about Questions. Inquiries into the Cognitive Bases of Surveys. (Russell Sage Foundation, New York.)

Tourangeau, R., and K.A. Rasinksi, 1988, Cognitive Processes underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol. 103, No. 3, pp. 299-314.)

Appendix 6.1. CAQI Questionnaire for income questions used for in-depth interviews

Think-aloud instruction with questions on education level.

Aloud

```
>> LAB: THINK-ALOUD question:
        Arbitrary order of thinking-aloud and answering.
Until now I have asked you to answer the question straightforward.
With the following questions, it will be a little different.
Would you give an answer as before, but please THINK-ALOUD and tell me
everything that occurs to you? Hence you MENTION EVERYTHING that comes to
your mind.
I would like to hear from you how you react to the questions,
i.e. YOU FOR YOUR PART.
Please, tell me also what you think while looking for an answer.
It is up to you to decide when you give the answer.
What really MATTERS is that you THINK-ALOUD.

>> LAB: NEUTRAL PROBES: - Tell me what you think? I'm listening.
                        - What are you thinking now?
                        - Is there anything more?
                        - How did you get to this answer?
```

Part 1. Co-ordinated income questions

IntroInc

```
The following questions are about (your income / the income situation in
this household).
For a survey like this, it is important to be able to roughly cluster
households in income groups.
```

HdBreadW

```
>> LAB: Lab question: What do you understand by ***>

In the next question I will be talking about the head breadwinner.
With that I mean:
***> THE PERSON IN THIS HOUSEHOLD WITH THE HIGHEST NET INCOME.

>> lab: - Who is the person with the highest net income ?
        - Check whether this is easy to determine by respondent?

OBSERVE: Is respondent the head breadwinner?

1. yes,
2. no
```

InSource

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

>> INT: - Major income source = income source from which
          the head breadwinner receives the highest net income.
        - Income from alimony, riches, dividend and free-lance
          work should be listed as income 'from another source'
          (= code 6). <<

From what source (do you / does the head breadwinner) receive the highest
net income?

Is that: -->
1. from wages or salary,
2. from profit from own business,
3. from early retirement,
4. from pension scheme, old age benefit, widow benefit,
5. from social benefit, half-pay or scholarship,
6. from another source,
7. refuses,
8. doesn't know.
```

NoNetInc (to be asked in case of proxy reporting: HdBreadW = No)

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

Do you know the net income from (you/this household), approximately?

>> INT: This concerns the income after taxes and insurance
        premiums have been paid.

1. yes,
2. yes, but refuses,
3. no,
4. doesn't know
```

YearIncS (self-employed: InSource = profit or InSource = another source)

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

This question is about the net income, so the income after taxes and social
charges have been paid.
You do not have to count money for childcare, in case you receive some.
Could you, rounded in thousands of guilders, tell what the net income for
(you/this household) has been in ***> THE LAST 12 MONTHS?
(In case more persons have an income of their own, could you add all
incomes together?)

>> INT: - The net income rounded in 1000 guilders
        - If necessary (an estimation of) the income in 1993 will do. <<
```

MnthInc

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

This question is about the net income in an ordinary month, so the income
after taxes and social charges have been paid.
You do not have to count money for childcare or vacation benefit, in case
you receive some.
Could you, rounded in hundreds of guilders, tell what the net income for
(you/this household) is in ***> AN ORDINARY MONTH?
(In case more persons have an income of their own, could you add all
incomes together?)

>> INT: - The net income is the income paid into the bank or
          postal account in an ordinary month.
        - In case of very fluctuating earnings e.g.
          with irregular jobs or hours of labour, answer 'doesn't know' <<
```

YearInc (MnthInc = doesn't know)

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

Could you, rounded in thousands of guilders, tell what the net income for
(you/this household) has been in ***> THE LAST 12 MONTHS?
Again, this is about the net income, so the income after taxes and social
charges have been paid.
You do not have to count money for childcare, in case you receive some.
(In case more persons have an income of their own, could you add all
incomes together?)

>> INT: - The net income rounded in 1000 guilders
        - If necessary (an estimation of) the income in 1993 will do. <<
```

Answer (additional lab question)

```
>> LAB: Lab question

>> LAB: IF NECESSARY
        - Probe how the respondent came to his answer:
          calculations.
        - respondent refers in his/her answer to:
          1. own memory
          2. housemate or other person (e.g. accountant)
          3. bank account slip / postal account slip
          4. payment slip / wage slip / pay-roll
          5. tax form
          6. other administration / other source, being: ....
>> LAB: Enter a short description or the corresponding number.
```

Accurate (additional lab question)

```
>> LAB: Lab question

>> LAB: Ask for accuracy: What does respondent think about it:
        - How accurate do you think the income you mentioned, is?
        - How accurate do you think the income you mentioned, is when
          expressed in guilders?
        - How accurate do you think your answer is?
        - How many guilders does the income mentioned by you,
          differ from your actual income, do you think ?

 1.      1 guilder
 2.      5 guilders
 3.     10 guilders
 4.     25 guilders
 5.     50 guilders
 6.    100 guilders
 7.    250 guilders
 8.    500 guilders
 9.   1000 guilders
10. > 1000 guilders
```

NumInc

```
OBSERVE: Is there one income or more?

1. one income
2. more than one income
```

HlthIns

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

Do you still have to pay health insurance premium from this income?

1. yes
2. no
```

TxReturn

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

Do you expect to get a tax return ***> ON THIS YEAR?

1. yes
2. no
```

HowMuchR

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

Could you tell how much, approximately?
```

Part 2. Additional test question (fill-in scheme)

IntroAdd

```
Now, I would like you to go through the income question once more.
We will do that in another way as just now.
I would like to know the best way to ask for income data.
Therefore, I hand over this fill-in scheme to you.

>> LAB: Hand over fill-in scheme including pen, now.

Would you be so kind to answer the income question once more,
but now by means of this scheme.
I would like to hear from you what you think of this scheme.
I read the question to you once more.

>> LAB: Press <1> for question.
>> LAB: Respondent him/herself determines how to handle, but:
        - in the first column: write the names of
          all housemates who have an income of their own.
        - in the appropriate corresponding columns: write income.
        Respondent does not need to make calculations.
        Enter amounts into the computer.
```

Fill-in scheme (1) to determine the net household income in an ordinary month:

Who in your household have an income of their own?	How much net income is this in an ordinary month, from:					
	wages or salary	profit from own business	early retirement	pension scheme, old age benefit, widow benefit	social benefit, half pay, or scholarship	another source, being: ……
	You do not have to count money for childcare or vacation benefit, in case you receive some.					
	f	*f*	*f*	*f*	*f*	*f*

Fill-in scheme (2) to determine the net household income in the last 12 months:

Who in your household have an income of their own?	How much net income is this in an ordinary month, from:					
	wages or salary	profit from own business	early retirement	pension scheme, old age benefit, widow benefit	social benefit, half pay, or scholar-ship	another source, being: ……
	You do not have to count money for child care, in case you receive some.					
	f	*f*	*f*	*f*	*f*	*f*

YearInS2 (self-employed: InSource = profit or InSource = another source)

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

This question is about the net income, so the income after taxes and social
charges have been paid.
You do not have to count money for childcare, in case you receive some.
Could you, rounded in thousands of guilders, tell what the net income for
(you/this household) has been in the last 12 months?
(In case more persons have an income of their own, could you add all
incomes together?)

>> INT: - The net income rounded in 1000 guilders
        - If necessary (an estimation of) the income in 1993 will do. <<
```

MnthInc2

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

This question is about the net income in an ordinary month, so the income
after taxes and social charges have been paid.
You do not have to count money for childcare or vacation benefit, in case
you receive some.
Could you, rounded in hundreds of guilders, tell what the net income for
(you/this household) is in an ordinary month?
(In case more persons have an income of their own, could you add all
incomes together?)

>> INT: - The net income is the income paid into the bank or
          postal account in an ordinary month.
        - In case of very fluctuating earnings e.g.
          with irregular jobs or hours of labour, answer 'doesn't know' <<
```

YearInc2 (MnthInc = doesn't know)

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

Could you, rounded in thousands of guilders, tell what the net income for
(you/this household) has been in the last 12 months?
Again, this is about the net income, so the income after taxes and social
charges have been paid.
You do not have to count money for childcare, in case you receive some.
(In case more persons have an income of their own, could you add all
incomes together?)

>> INT: - The net income rounded in 1000 guilders
        - If necessary (an estimation of) the income in 1993 will do. <<
```

Scheme

```
>> LAB: THINK-ALOUD question: No order of thinking and answering
        NEUTRAL PROBES: Tell me what you think? I'm listening.

>> LAB: Probe how respondent got to the numbers filled in.
>> LAB: Respondent him/herself determines how to handle, but:
        - in the first column: write the names of
          all housemates who have an income of their own.
        - in the appropriate corresponding columns: write income.
        Respondent does not need to make calculations.
        Enter amounts into the computer.
```

Net (additional lab question)

```
>> LAB: MEANING ORIENTED PROBE: What do you understand by ***>

In these questions we spoke about ***> THE NET INCOME.

>> LAB: Ask what respondents understands by ***> THE NET INCOME.
        - What is according to you THE NET INCOME?
        - Could you describe this?
        - What do you understand by THE NET INCOME?
```

Sensitiv (additional lab question)

```
>> LAB: TEST QUESTION on sensitivity of income

You did report your income.

Could you tell me in what situations you will report your income and in
what situations you won't?
```

Chapter 7

Case Study 2: Testing Questions from the European Community Household Panel

Pre-tests on ECHP Questions on Daily Activities, Pension Schemes, and Education and Training

Summary: *This chapter presents the results of a laboratory pre-test study on a number of questions from the European Community Household Panel (ECHP), a Eurostat survey. These questions asked for daily activities, pension schemes, vocational education and training. In cognitive interviews (using thinking aloud with follow-up probes, and meaning-oriented probing in a CAQI tool) the wording of the questions was tested. For vocational education and training also vignettes were applied to estimate the degree of ambiguity on these issues. The results are used for suggesting improvements in the questionnaire.*

This pre-test study has been carried out by order of the former Division of Socio-Economic Statistics of Statistics Netherlands, in order to advice Eurostat on the quality of the ECHP questionnaire. The pre-test results have been presented to the client (the Division of Socio-Economic Statistics) by Snijkers (1995a, 1995b). Based on these reports, Van de Donk and Snijkers (1995) reported to Eurostat. Eurostat (1995) reported back that refinements were proposed for the questionnaire, making use of the "excellent evaluation study of selected questions carried out by the Netherlands CBS for Eurostat." This chapter is a slightly adapted version of Snijkers (1995b), and provides a second example of a laboratory pre-test study and how pre-test results are presented to the client. In contrast to chapter 6, here I will concentrate on suggestions for improvement of the questions (within the ECHP layout), since the client was mainly interested in these recommendations and not in a detailed description of the identified problems in the questions or the design of the pre-test study.

Keywords: *Pre-testing Questionnaires, Cognitive Interviews, Daily Activities, Pension Schemes, Vocational Education or Training*

7.1. Introduction

This report discusses the results of pre-tests on questions in the 1995 wave of the ECHP questionnaire (European Community Household Panel, an Eurostat survey). It was suspected that several of these questions might cause problems in the field, impairing data validity. The pre-tests were carried out by the Questionnaire Laboratory of the Department of Data Collection Methodology at Statistics Netherlands in 1995.

About 20 ECHP questions were evaluated by laboratory research. These are questions on daily activities (Individual Questionnaire, Qs 1-3, 65), on contributions to pension schemes (Qs 36-44) and on vocational education or training (Qs 121-129). (See appendix 7.1 for the question wording.)

These questions were pre-tested by means of cognitive interviews. Cognitive interviews indicate what kind of problems might occur in the question-and-answer process with these questions, in particular with adequate understanding and answering of the questions (comprehension and interpretation). This method can be said to confront 'the concept as measured' against 'the concept as intended'. Cognitive interviews can also indicate why these problems arise, and how the questions could be reformulated to avoid these problems: From the results one can gather why misunderstandings arise, and in what direction a solution might be found.

Since the questions were intended to be posed to a respondent by an interviewer, the cognitive interviews were also face-to-face interviews. In the cognitive interviews respondents were asked to think aloud: to tell their thoughts about what comes to their mind while answering. Also, they were probed on several terms used in the questions, i.e. they were asked how they interpreted these terms in the context of the question. To get an impression of their spontaneous ideas, the closed questions were read aloud first without answering categories. Whenever necessary these categories were read to the respondents to get a proper answer. In this way an impression can be got about quality of the categories and about their influence on the answers.

On one item, vocational education, also another method of detecting sources of possible flaws was used during the cognitive interviews, the so-called vignette method. This could be called determining 'the spontaneous concept'. Fourteen vignettes (situations of vocational education and training) were read to the respondents. They were asked to classify them as vocational education, training or something else. This method clarifies both the borderlines between the different situations and its potential variation in the perception of the respondents.

Originally the questions were formulated in the English language. In order to conduct the cognitive interviews with Dutch respondents, these questions were translated into Dutch. They were translated as literally as possible, with due allowance for the Dutch context. With these

questions 24 independent cognitive interviews were carried out[1] with carefully and purposively selected volunteering respondents. The 24 respondents have been selected so as to distribute as evenly as possible over education level (lower and higher level), gender, and age (20-<30 and 30-<50 years); and furthermore, they had to work in a job (either for less or for more than 15 hours a week) and attend (or have attended recently) some kind of course. In appendix 7.2 and 7.3 the characteristics of the selected respondents are presented.

In a preliminary pre-test session, 8 cognitive interviews have been carried out with the same questions, to get a first impression of the problems with the questions (Snijkers, 1995a). These 8 cognitive interviews seemed to indicate that it is hard for people to answer questions on pension schemes, since they do not know much about pensions. And also, the respondents who volunteered for these 8 interviews did not seem to classify the same kind of training or course in the same way. But, due to the very limited number of interviews and breakdowns by routing in the questionnaire, we concluded that additional and more detailed cognitive interviews were needed.

In the next section general conclusions and suggestions for improvement are stated for the three topics in the ECHP questions: daily activities, pension schemes and vocational education. These conclusions are based on the 24 cognitive interviews, including the vignette study. In section 7.3, detailed results will be described. Whenever appropriate, some supporting evidence is given, including descriptions of respondents' reactions and our interpretation thereof. Other arguments (continuity, time-use, survey planning) will have to be weighted against the expected quality gain of the suggested improvements.

7.2. Conclusions and suggestions

In this section the general conclusions on the three researched topics of the ECHP questionnaire are stated: in subsection 7.2.1 for the questions on daily activities, in 7.2.2 for those on pension schemes, and in 7.2.3 for those on vocational education or training.

7.2.1. Daily activities (Qs 1-3, 65)

The questions on daily activities did not seem to raise major problems for the respondents, except for **Q3** (on the kind of job).

[1] The interviews were carried out by Questionnaire Laboratory researchers: Hans van Kerkoerle, Math Bosch, Annemieke Luiten and the author.

To sum up our findings on the questions on daily activities, we can say that **Q1** (on working in a job of at least 15 hours a week) was well comprehended. 'Working in a job or business' was interpreted roughly in accordance with the definition. This concept was not ambiguous for the respondents, with exception of 'military service' and '*unpaid* work in a family enterprise'. Therefore, we suggest highlighting these remarks in the questionnaire.

There seems to be overlap between **Q2** (temporarily absent from job) and **Q1**. To solve this problem, two solutions seem possible: to combine the two into one question or to change the wording of **Q2**. The combined question could be formulated as:

Q1' *Do you have a job or business normally involving at least 15 hours of work a week, including paid and unpaid work in a family enterprise, and also including a job from which you are temporarily absent for some reason?*

Yes ( ) → *Q3*
No ( ) → *Q65*

Or the wording of **Q2** could be changed into (this question will only be posed if the answer to **Q1** is 'No'):

Q2' *Is this because:*
- *You simply do not have a job of at least 15 hours of work a week.* ( ) → *Q65*
- *You have a job of at least 15 hours a week, but from which you are temporarily absent for some reason.* …… ( ) → *Q3*

Q3 (on what kind of job) is too long to comprehend immediately. And the answering categories seem not to be mutually exclusive. So, respondents should be allowed to choose more than one option or they should pick the main activity. Therefore, we suggest splitting this question up into two questions, asking about the basic options first:

Q3'a *Are you:*
- *Working with an employer,* …… ( ) → *Q3'b*
- *Self-employed, or* …… ( ) → *next*
- *Working in a family enterprise.* …… ( ) → *next*

Q3'b *Are you working with an employer:*

(Tick all that apply) or *(If more, tick main activity)*
- *In paid employment,* …… ( ) → *next*
- *In training under special scheme related to employment,* …… ( ) → *next*
- *In paid apprenticeship.* …… ( ) → *next*

Respondents do not seem to have any difficulties in comprehending question **Q65** (on main activity status), but it is difficult for respondents to come up with an answer that fits into one of the answering categories. From the 8 preliminary interviews we had done (Snijkers, 1995a), we also know that the list of categories is incomplete: the category 'in paid employment for less than 15

hours a week' is lacking. Therefore, we suggest putting the answering categories on card, including 'in paid employment for less than 15 hours a week'.

7.2.2. Pension schemes (Qs 36-44)

As for 'pension schemes' it appeared that most respondents do not know much about this topic. After the introduction or after the first question, almost all respondents replied with: "I have no idea about how my pension is arranged!"; "I don't know much about that!" This indicates that the answers one gets on these questions may not be very reliable. Therefore, we suggest leaving these questions out of the questionnaire, with the exception of the questions on 'job-related pension scheme' (**Q36**) and 'private pension scheme' (**Q43**). From the cognitive interviews it seems that people are able to answer these two questions for the present situation, but one should not go into technical details or ask about the past.

This conclusion should not be generalised to other countries of the Union without further investigation. The situation that causes the lack of knowledge might be specific for the Netherlands. The majority of employees in the Netherlands are in an occupational pension scheme, mostly with a mandatory character, and this has been the situation for a long time. As practically no action or decision is expected from the employee little or no knowledge is required with respect to pension schemes. The answers to the pension questions show that this is especially the case with financial details and information from the past. Therefore we suggest reconsidering the questions on pension schemes with the aim to get a better (and maybe a more complete) picture which can be expected to be comparable between countries. It should be checked whether such situations exist in other Member States of the Union.

The questions **Q37-Q41** ask technical details about the job-related pension schemes. As stated above, we suggest leaving these questions out of the questionnaire to be used in the Netherlands. An alternative option might be to change the wording of **Q37**, and to combine the questions **Q38-Q41** into one question:

Q37' *Who contributes to this pension scheme?*

Fully by employer …….. *( ) → next*
Both myself and my employer …….. *( ) → Q38'*
Fully by oneself (directly or as deduction) …….. *( ) → Q38'*

Q38' *How do you pay for this pension?*

(Tick all that apply) (Answering categories on card)
Deduction from wage or salary by the employer …….. *( ) → next*
Mandatory additional contribution by yourself …….. *( ) → next*
Voluntary additional contribution by yourself …….. *( ) → next*

As for the past, from the cognitive interviews it became apparent that **Q42** (pension entitlement from employment with past employers) was hard to answer as well. Therefore, we also suggest leaving this question out. As an alternative option, we suggest to add a filter question on past employment: Do you have had past employers?

With regard to the question on private pension schemes (**Q43**), it became clear that the respondents know what arrangements they have made themselves. A problem might be that people do not report this with **Q43**, because they do not see their arrangement as a private pension scheme. To overcome this problem, we suggest changing the question into:

Q43' *Apart from the job-related pension scheme, do you have made yourself any arrangements for your pension (or old age)?*

Yes *( )*
No *( )*

7.2.3. Vocational education and training (Qs 121-129)

The questions on vocational education or training were in general well comprehended by the respondents. The questions **Q121** (vocational education or training), **Q122** (duration), **Q123** (starting and finishing dates), **Q125** (payment and organisation of course by employer) and **Q128** (usefulness) did not seem to cause considerable trouble. With the other questions: **Q124** (level and type of course), **Q126** (kind of administration of course), **Q127** (reason for taking the course) and **Q129** (other course), we detected some problems.

Q124 (level and type of course) was very difficult for the respondents to answer. Respondents did not know how to classify the course or training they were talking about. Especially, the answering categories gave rise to a lot of hesitation: the respondents could not remember all categories at once (so the categories had to be repeated), and it was hard for them to choose an appropriate category. As a result, the respondents did not classify the same kind of course or training in the same way. Some respondents classified a short course as 'training at a vocational school', while others (who had attended the same short course) finally came up with 'training in a working environment'. Thus, the data on this question may not be reliable.

These reactions indicate that the category 'short courses given outside the working environment but aimed at doing the job' should be added to the list. Another solution is to select answering categories only according to: systems providing work experience, instructions elsewhere, or both. This would also overcome another problem with the answering categories: because categories on level and on type of education are listed in the same list, there seems to be an overlap between the categories. In any case, to make the question easier to comprehend, we suggest putting the answering categories on card.

The difficulty in classifying activities as training, education or otherwise did show clearly from the application of the *vignettes*. The respondents often classified specific situations differently. Some situations were very clear, especially those having to do with vocational education. The respondents consider a situation to be 'vocational education' when it is an education for a longer period of time aimed at learning a specific profession. Situations having to do with 'off the job training' (short courses) were mostly considered to be training. But there is a broad field of situations that are not clearly considered as training. These situations range from 'on the job training' in the sense of supervision to 'informal activities' like talking to co-workers, reading, or education for pleasure. Therefore, we recommend assuring that the respondents understand 'training' as intended (e.g. by adding an interviewer instruction on unstructured training), and excluding explicitly extremely short courses, and/or normal activities to keep up with developments in the field of work, like reading technical articles.

Although **Q125** (payment and organisation of course by employer) is a double-barrelled question, it did not cause difficulties to the respondents. However problems may arise in certain circumstances, like when the employer paid for the course but did not organise it. Therefore, the question could be reformulated. Firstly, the question might be stated in the active voice, which is easier to understand than the passive voice, and secondly, the answering categories could be specified as follows:

Q125' *Did your employer pay for or organise the course?*

Yes, employer only paid for course …… *( )*
Yes, employer only organised course …… *( )*
Yes, both …… *( )*
No, none of these …… *( )*

As for **Q126** (full-time, part-time, course by correspondence) it appeared that the list of answering categories is not complete: the answering category 'short course' (a course of a few days or less) is lacking.

In **Q127** (reason for taking the course) two reasons are mentioned. Some respondents had difficulties in recalling the reasons. While answering this question, they only recalled the last reason. Therefore, we suggest changing the wording of the question:

Q127' *For what reason did you take the course?*

(Tick all that apply)
To improve skills …… *( )*
Job prospects …… *( )*
Other reason: _____
No reason at all …… *( )*

The last ECHP question posed to the respondents was **Q129** on attending a language course or some other adult course. From the respondent reactions it became clear that the answering category 'no' is lacking (see also Snijkers, 1995a). Also, there seems to be no clear distinction with **Q121**, due to ambiguity in the interpretation of 'adult education course'. Therefore, we suggest changing **Q129** into the following questions:

Q129'a *Apart form any course already mentioned, have you at any time since January 1994 attended some other course?*

Yes ( ) →*Q129'b*
No ( ) → *next*

Q129'b *What kind of course was that?*
_____

7.3. Results of the cognitive interviews

In this section the results of the 24 cognitive interviews on which the conclusions are based, will be described: in subsection 7.3.1 for the questions on daily activities, in 7.3.2 for those on pension schemes, and in 7.3.3 for those on vocational education or training.

7.3.1. Daily activities (Qs 1-3, 65)

Q1: Working in a job of at least 15 hours a week
As for **Q1**, the respondents reacted immediately saying whether they had a job or not. People who are working in a full time job do not have any trouble at all to answer this question. This question might raise some difficulties for people who have two or more part-time jobs, or people who are working irregularly. They have to sum up all the hours they make, or calculate an average. But this did not seem to be very troublesome for the respondents. As one respondent said: "No, at the moment. It varies from week to week. Some weeks I work for more than 15 hours. I am a stand-in teacher at primary school. But, not at the moment, no!"

The respondents were also probed for their interpretation of 'working in a job or business'. With regard to this concept, the respondents referred to "having regular work", "working with an employer on basis of a contract", "doing work, in exchange for which you get paid". Most respondents also considered 'any paid training under special scheme related to employment, or block release or other paid apprenticeship' to be 'working in a job or business'. Those who did not consider this to be 'working in a job' said that "this is improving *your own* skills; basically you do it for yourself". And furthermore they replied that "this is not working"; "it is no labour"; "you do not do what you are hired for to do". 'Paid work in a family enterprise' is also considered to be 'working in a job', but 'unpaid work' is not, because you are not paid for the work you do. As for

'Military service' 12 respondents said that this is also 'working in a job', while the other 12 respondents did not think so. Those who said that 'military service' is also 'working in a job', thought so because you are hired to do a job, whether it is mandatory or not.

Q2: Temporarily absent from job

As stated in subsection 7.2.1, there seems to be an overlap between **Q1** and **Q2**. **Q2** was posed to 10 respondents (those who answered 'No' to **Q1**). Some of them reacted somewhat irritated to this question saying that they had already said that they do not have a job: "As I have said before, I work irregularly. The answer is no."

Q3: What kind of job

From the interviews it appeared that **Q3** is too long to comprehend immediately. This question was posed to 14 respondents (those who answered 'Yes' to **Q1**). For 7 respondents the question had to be repeated, indicating that they could not comprehend the question at once. They said: "To me the question is not clear, it's too long. Could you repeat the question?" After the question was repeated, they immediately responded.

One respondent gave a double answer to **Q3**. He said "the first two categories", which are 'in paid employment' and 'in training under special scheme related to employment'. This respondent had to attend training courses for his job on a regular basis. This indicates that, in accordance with the definition, the answering categories are not mutually exclusive.

Q65: Main activity status

Q65 was posed to 10 respondents (those who answered 'No' to **Q2**). After the question was read aloud, the respondents started to talk about their 'main activity status', in the sense of what they do "most of the time", i.e. "the activity most of the time spent on", "the activity that takes the most hours". One respondent came up with a rather different interpretation of 'main activity'. He said: "My main activity is the activity which is the most important, prominent to me." To 7 respondents it was not immediately clear what kind of answer was expected from them. So, the answering categories had to be read aloud. After that, they came up with an answer immediately. This indicates that the meaning of the question is not clear to these respondents without knowing the answering categories.

7.3.2. Pension schemes (Qs 36-44)

Q36: Job-related pension schemes

Q36 was posed to all 24 respondents. Fifteen respondents almost immediately came up with an answer (yes or no). Some of them replied: "Yes, I thought so. It is indicated on the pay-slip. But this is all I know about pensions." Those who hesitated replied: "I don't know. It is a good

question, but I really don't know. But, I do remember that there is something on my pay-slip about pensions. So, let's say 'yes'."

From the interpretations of 'job-related or occupational pension scheme' it became clear that the respondents know what the question is about. This concept is interpreted as: "The pension scheme that is arranged by your employer. He takes care of the payment to the pension fund. The premium is paid automatically."

Q37: 'How is your pension paid for?'

Q37 was posed to 17 respondents (those who said 'yes' to **Q36**). Most of them knew that some part of their wages goes to the pension fund, but they did not know whether the employer also pays his share. Because they were not sure, 5 respondents said that their pension was fully paid for by themselves, and 3 answered "don't know". In Dutch reality, in almost all pension schemes the employer pays more than the employee does, and sometimes all.

From the reactions it also became apparent that 'to pay' (in the Dutch translation) has two interpretations: one regards ownership ("It's my money. So essentially, I pay for my pension."), and the other is 'to wire the money' ('the activity of transferring the money from one account to another'). Because of this last interpretation one respondent answered with 'fully paid for by the employer'. The other 8 respondents hesitated and finally answered "both".

Apart from the fact that the respondents do not have the knowledge to answer this question, they interpreted the question differently than intended. They spontaneously said: "How? I thought that it is automatically paid for by my employer, by deduction from my salary." (This is the second interpretation of 'to pay'.) Then the answering categories were read aloud and the respondents said: "Oh, now I know what you mean. I pay for it, but I don't know about my employer." Their spontaneous reactions more or less are an answer to **Q38**, on direct contributions to the pension scheme. So, as we have seen with **Q65** (in subsection 7.3.1), here again we see that the interpretation of the question is dependent on the answering categories.

Q38: Direct contributions to the pension scheme

Q38 was answered with 'no' by all respondents, since they had indicated that their payment to the pension scheme was automatically done by their employer. But for some respondents this question was difficult to comprehend, due to the exclusion 'other than', which may result in a double negation: one in the question and one in the answering categories. One respondent said: "The question was whether it is *not* arranged in that way. So, whether the payments to my pension scheme is arranged differently than by deduction from my salary. In that case, the answer is 'no': no direct contribution." So he first had to paraphrase the question to understand its meaning.

The meaning of 'direct' is dependent on the point of view. In **Q38** 'direct contribution' means "any payment by you, without intervention of the employer". However, from spontaneous reactions

of the respondents it became clear that 'direct' to the respondents means "that I do not have to worry about it. It is automatically paid for by my employer, directly to the pension fund." This is the opposite point of view. However, this did not seem to make the answer less valid.

Q40: Voluntary contributions

Q40 is not applicable for the Dutch situation. However, we put the question into the interview to test it. All respondents answered correctly with 'no' (which means exclusion of **Q41**). The interpretation of 'voluntary' was: "that I myself have the choice to make any additional contribution to the job-related pension scheme, and that I can decide how much I pay every month. So, this arrangement is not mandatory."

Q42: Pension entitlements from past employment

The question was well comprehended, but most respondents who have had past employment were not sure about pension entitlements arising from that employment (**Q42**). They mostly replied something like: "I don't know. I guess so, but I'm not sure. I guess there had also been some deductions from my salary. So, let's say 'yes'." For respondents who have had no past employers, this question was not difficult to answer. They immediately replied with 'no'. However, this indicates that that the 'no' answer to this question has two meanings: 'No, I do not have pension entitlements arising from past employment, although I have had past employers', or 'No, I do not have such pension entitlements, since I do not have had past employers'.

Q43: Private pension schemes

The question on private pension schemes (**Q43**) was well comprehended by most respondents. 'Private pension scheme' was interpreted as "a pension scheme arranged and fully paid for by me". The respondents made it clear that they know whether they had made any arrangements themselves. They replied with: "Whether I have made such arrangements myself? No!", or "Yes, I have an extra insurance for later." A few respondents, however, were not sure about their life insurance[2]. They said: "In case a life-insurance is a pension scheme, I say 'yes'." It is clear that respondents know what arrangements they have made themselves, but they might not report this with **Q43**, because they (probably rather often) do not see their arrangements as a private pension scheme, but merely as long term savings.

[2] Many but not all life insurances accumulate capital. This is paid back together with accumulated interest, on a fixed age, e.g. 65 or 70. (In a pension scheme capital is paid back by regular amounts during the remaining lifetime.) The reason to pay for this kind of life insurance is of course often providence for old age. In the Netherlands large amounts are saved in this way. The questionnaire interpretation does not discuss this item at the moment.

7.3.3. Vocational education and training

7.3.3.1. Questions on vocational education and training (Qs 121-129)

Q121: Having been in vocational education or training

The part in the questionnaire on vocational education and training started with **Q121**. This question was rather well comprehended by the respondents. 'Vocational education' was interpreted as "specific education for a longer period of time, aimed at learning a profession". Whereas 'training' was interpreted as "a short course, following vocational education, aimed at learning new skills in conjunction with your profession" or "a short course aimed at rubbing up your skills". Some respondents had other spontaneous associations like "sports training". For these respondents, the use of the word 'training' in conjunction with 'vocational education' at first appeared odd to them, but when they came to think about it, they came up with the interpretation listed above.

Q122 and Q123: Duration, and starting and finishing dates

It was not very troublesome for the respondents to recall events connected with 'vocational education' and 'training' over a period of 16 months (from January 1994 up to April 1995).

Q124: Type of vocational education or training

In contrast to **Q121**, **Q124** was troublesome. The respondents had great difficulty to answer this open question. They spontaneously replied with a description of the course or training, like: "It was a course at the Open University", "It was a training in how to answer the phone and how to deal with people", or "It was a course in client friendliness". Then, the answering categories were read aloud. However, the respondents could not remember all options (7 in the Dutch question). So, the options had to be read twice.

In case of 'third level vocational education', there was no problem to give an answer, once the answering categories had been read aloud. However, the three categories on 'specific vocational training' were not easy to interpret and the list of categories does not seem to be complete and prone to errors. The respondents who had attended a short course in relation to their work did not know what to choose. They hesitated and then said something like: "It was a training given somewhere else, but it was connected to my work. So, I would say 'at a vocational school' and 'in my working environment', although 'vocational school' is not the right word. It was given at some institution where you can take all kind of courses." What they meant was that the training, which was not given at a 'vocational school' or in a 'dual system', was aimed *at* the working environment (to be better trained to do the job[3]) and *not* that the training was given *in* the working environment. So, the category 'short courses given outside the working environment but aimed at doing the job' should be added to the list.

Furthermore, it appeared that two dimensions of education and training are put together in the answering categories: level of education and type of education. To illustrate this, one respondent said: "It was a course at an institution where you learn how to run a pub. It lasted for three months. I had to go there one evening every week. Let's say it is 'third level education', but I'm not sure." The level of this vocational education could very well be equivalent to a 'third level education', but this type of education could also be classified as 'specific vocational training at a vocational school'.

Q126: Full-time, part-time, course by correspondence

Here also the answering category 'short course' was felt lacking. Respondents who had attended a 1-day course answered with: "It was not part-time, and it was not by correspondence. So, full-time: I was there one whole day." Another respondent said: "It was a 1-day course, from 9 to 5. This is not full-time, because I did not go there for a longer period of time during daytime. It also is not part-time. That is when you e.g. attend a course during the evening for a longer period of time, say once a week for half a year. And it is also not by correspondence. So, my course doesn't fit into the three categories listed."

Q127: Reasons for attending the course

Q127 is a question with a list of reasons for attending the course. Some respondents had difficulties in recalling the reasons that were read aloud. They only recalled the second part of the question, saying: "No, I did not attend the course for job prospects. The course is only useful in this job. I cannot imagine that another employer would hire me, because I did this course." From the conversation earlier on in the interview however, it was clear that these respondents had attended this course to improve their skills. So, the question was repeated. Then they replied: "Well, in that case I have to say 'yes': I took the course to improve my skills." So, we recommend changing the question wording.

Q129: Attending a language course or some other adult course

The last ECHP question posed to the respondents was **Q129**. There seems to be no clear distinction with **Q121**, due to ambiguity in the interpretation of 'adult education course'. As for this overlap with **Q121**, some respondents said: "Well, I consider the course we just talked about as an adult education course." This was a study at the Open University, a computer course in Word Perfect or a language course in English. These courses had been attended in relation to the job. They interpreted 'adult education course' as "a course for grown ups, being 18 years of age or older". The other respondents interpreted 'adult education course' as "a course for people who haven't been to school for several years and now, on a voluntary basis, attend a course for hobby, for general education or to learn for a new profession."

---

[3] It should be stressed that the number of this kind of training is large and will probably continue to grow, due to the need of permanent education.

One respondent remarked about the wording of **Q129**: “What an odd formulation. I would say: ‘Have you attended some other adult education course, like a language course?’ And besides, I would not talk about ‘other *adult* education course’ but simply about ‘other education course’.” Therefore, we recommend changing the question wording.

7.3.3.2. Vignettes

At the end of the cognitive interviews, 14 vignettes on vocational education and training were read to the respondents. The respondents were expected to tell whether a specific situation was to be considered vocational education, training or something else. Ten of these vignettes were copied from a SCPR study (Social Community Planning Research) on training (Campanelli & Channell, 1994). Apart from these vignettes, four vignettes were added with regard to attending short courses and apprenticeship. In table 7.1, the results of our experiment are given.

These vignettes show that people do not classify a specific situation with regard to vocational education and training in the same way. On the one hand, some situations are very clear, especially those having to do with vocational education. The respondents consider a situation to be ‘vocational education’ when it is an education for a longer period of time aimed at learning a specific profession. Situations having to do with ‘off the job training’ are mostly considered to be training. These situations are characterised as a short course with a clear starting and finishing point, given by someone outside the working environment who is able to tell what is wrong and what is right, and aimed at training skills. On the other hand, there is a broad field of situations that are not clearly considered as training. These situations range from ‘on the job training’ up to ‘informal activities’ like talking to co-workers, helping, training yourself, self-development or hobby.

Table 7.1. Vignettes on 'vocational education and training'

Vignette (*)	Situation vignette is supposed to indicate	Respondent interpretation (number of respondents)		
		Vocational education	Training	Something else, namely:
When George started in a new job a more experienced worker would sometimes give him hints on how to do the work. (**)	Informal activity: talking to co-worker	2	6	16 supervision helping
Susan reads articles in technical journals in the evening to help her keep up with new developments in her field of work. (**)	Informal activity: reading	1	6	17 self-development keeping up to date
Ellen works as a supervisor. She says she acquired her management skills in the 'school of life'. (**)	Informal activity: 'school of life'	1	5	18 self-development own experience
As Graham was new to the factory, he did not always know how to set up the machinery to meet a new order, but when this happened, his supervisor showed him what to do. (**)	On the job training	1	8	15 supervision explaining helping
Fred says he now knows how to do his job because he has learned from his mistakes. (**)	Training vs. job experience	3	7	14 own experience
John works for a land-clearing firm. His boss sent him to a day to an equipment manufacturer to learn how to use a chain saw safely. (**)	Off the job training	5	17	2
James pays to take evening classes in Spanish because he enjoys learning languages. (**)	Education for pleasure	4	2	18 hobby
After leaving school last year, Elizabeth is studying full-time to qualify as a medical doctor. (**)	Vocational education	24	-	-
Elizabeth's father is a family doctor. He attended a two-day course on the modern practice of a doctor.	Off the job training	4	17	3
Mark is a teacher in law at an economical college for more than 10 years now. For 6 months he accompanies a barrister to see what the law-practice is nowadays.	Apprenticeship	3	12	9 apprenticeship
Tom is a student in a 'dual system'. This means that every week he is attending classes for four days, and working with an employer for one day to get working experience.	Vocational education and apprenticeship	23		1
Michael has picked up his skills as a car mechanic by working on his friends' cars. (**)	Self-initiated activity	-	3	21 self-development hobby
Miranda does some typing at work and she goes to evening classes to improve her typing, even though they don't lead to a typing qualification. (**)	Off the job training	2	17	5 rubbing up skills
Robert is studying for aircraft constructor. He went to Fokker last year for a 6-months apprenticeship.	Vocational education and apprenticeship	20	3	1

(*) The vignettes are listed in the order they were read to the respondents.
(**) Campanelli & Channell, 1994.

References

Campanelli, P., and J. Channell, 1994, Training: An exploration of the word and the concept with an analysis of the implications for survey design. (The Survey Methods Centre, Social Community Planning Research, London.)

Eurostat, 1995, Working Group "European Community Household Panel": Questionnaire for Wave 3: proposed changes. (Eurostat report, doc.PAN 54/95. Statistical Office of the European Community, Luxemburg.)

Snijkers, G., 1995a, Pre-tests on ECHP questions on daily activities, pension schemes, and education and training: preliminary results. (CBS report, BPA no. H2117-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1995b, Pre-tests on ECHP questions on daily activities, pension schemes, and education and training: final results. CBS report, BPA no. H8170-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Van de Donk, P., and G. Snijkers, 1995, Testing on ECHP questions on daily activities, pension schemes, and education and training: final report. (CBS report, BPA no. H2876-95-SIP. Statistics Netherlands, Division of Socio-Economic Statistics, Department of Integration and Presentation, Heerlen.)

Appendix 7.1. ECHP questions

Individual questionnaire (Questions marked with '*' have been tested on request of Eurostat.)

1 We would like to start with your present work and daily activities. Are you at present working in a job or business normally involving at least 15 hours of work a week?

Definition (explain if necessary):
- "Job or business" includes any paid training under special scheme related to employment, or block release or other paid apprenticeship. It also includes paid or unpaid work in a family enterprise. Military/community service is not included.
- If you are working in more than one or business, please consider the total hours worked. If the total is 15 hours per week or more, please answer question 2 onwards in relation to your main job/business, i.e. the one with the most working hours.

Yes ( ) → 3
No ( ) → 2

2 Even if you are not working at present, do you have a job or business normally involving at least 15 hours of work a week, but from which you are temporarily absent for some reason?

Yes ( ) → 3
No ( ) → 65

3* Is this …

working with an employer
- in paid employment ..…..….. ( )
- in training under special scheme related to employment ……..... ( )
- in paid apprenticeship ……..…. ( )

self-employment ……..……...…. ( )
unpaid work in a family enterprise ……..…… ( )

65* What is your main activity status?[4]

In education or training …… ( )
Unemployed ……..…. ( )
Retired ……. ( )
Doing housework, looking after children or other persons ……..….. ( )
In community or military service ……. ( )
Other, please specify:
_____

[4] Please note that in this question 'status' is self-defined by the respondent.

Contributions to pension schemes

36 Now a few questions about pensions.

Are you a member of a job-related or occupational pension scheme?[5]

Yes	( ) →	37
No	( ) →	42

Please note that this refers to a job-related scheme, other than any general old-age pension scheme or purely private scheme.

37 How is your pension paid for?

Fully paid for by employer …….	( ) →	40
Jointly by myself and by employer ……..……	( ) →	38
Fully by myself (directly or as deduction) ……...	( ) →	38

38 Do you make any direct contributions to this pension, i.e. other than any deduction from your wages or salary by the employer?

Yes	( ) →	39
No	( ) →	40

39 How much do you contribute directly per month?

NC _____ per month

40 Do you make an additional voluntary contribution in conjunction with the employer's pension scheme in order to increase the amount of pension you will get?[6]

Yes	( ) →	41
No	( ) →	42

41 Does the employer make any contributions to this additional pension?[7]

Yes	( )
No	( )

[5] This does not refer to statutory pension schemes with universal coverage. This refers to a) all occupational or job related pension schemes, whether voluntary or mandatory, to which the employer, or the state in the case of public employees, makes a contribution; as well as b) occupational pension arranged by the person concerned, whether paid independently or through the employer. Both types may be supplemented by related but voluntary additions towards which the employer may or may not be obliged to contribute (Q40). In addition there are also purely private schemes (Q43).
[6] Insert country-specific term. This question may be dropped in countries where inapplicable.
[7] Insert country-specific term. This question may be dropped in countries where inapplicable.

42 Do you have any pension entitlement arising from employment with a past employer?

Yes ( )
No ( )

43 Do you contribute at present to a private pension scheme?

Yes ( )
No ( )

44 How much do you contribute per month to this private pension scheme?

NC _____ per month

Education and training

121* Have you at any time since January 1994 been in vocational education or training, including any part-time or short courses?

Yes ( ) → 122
No ( ) → 129

122 What is the overall duration of the course or training from start to finish?

Please answer in relation to any course currently being attended, otherwise in relation to the one that was finished most recently.

If less than 2 weeks: _____ days
If 2-9 weeks: _____ weeks
If longer: _____ months

123 What are the starting and finishing dates of the course?

Starting date		Finishing date	
1993 or earlier	( )	1994 month	_____
1994 month	_____	1995 month	_____
1995 month	_____	Still going on	( )

124* What type of vocational education or training course is/was it?

(Country-specific categories according to labour Force Survey)
Third level qualification, such as at technical college …….……. ( )
Specific vocational training:
- At a vocational school or college …… ( )
- Within a system providing both work experience and complementary instructions elsewhere
  (i.e., any form of 'dual system' including apprenticeship) …….…… ( )
- In a working environment
  (i.e., without complementary instruction at a school or college) …….…… ( )

125 Is/was the course paid for or organised by your employer?

Yes ……... ( )
No ……..... ( )
I was not in employment ….…… ( )

126 Is/was this a full-time attendance-course, a part-time attendance course, or a course by correspondence?
If part-time attendance course, please specify hours per week.

Attendance course full-time ……. ( )
Attendance course part-time ……..… hours per week _____
Course by correspondence ……..……. ( )

127 Was improving your skills or job prospects one of the reasons you took this course?

Yes ( ) → 128
No ( ) → 129

128 How useful do you feel this course has been for that purpose?

Very useful ……..……. ( )
Quite useful …….. ( )
Not so useful …… ( )
Not al all useful/waste of time ……. ( )

129* Have you at any time since January 1994 attended a language course or some other adult education course?

Tick all that apply
Language course ……..……. ( )
Other adult education ……... ( )

Appendix 7.2. Background characteristics of respondents with regard to selection criteria: gender, education level and age

Table 7.A2.1. Number of respondents by background characteristics: gender, education level and age

Age	Gender						Total
	male			female			
	Education level		Total	Education level		Total	
	low	high		low	high		
< 30	2	3	5	3	2	5	10
≥ 30	4	3	7	3	4	7	14
Total	6	6	12	6	6	12	24

Appendix 7.3. Background characteristics of respondents with regard to selection criteria and:
1. having a job of at least 15 hours
2. having a job-related pension scheme
3. attending a vocational education or training

Table 7.A3.1. Number of respondents with a job of at least 15 hours a week *), by background characteristics

Age	Gender						Total
	male			female			
	Education level		Total	Education level		Total	
	low	high		low	high		
< 30	1	3	4	1	-	1	5
≥ 30	4	1	5	2	2	4	9
Total	5	4	9	3	2	5	14

*) according to their own saying

Table 7.A3.2. Number of respondents with a job-related or occupational pension scheme *), by background characteristics.

Age	Gender						Total
	male			female			
	Education level		Total	Education level		Total	
	low	high		low	high		
< 30	1	2	3	-	1	1	4
≥ 30	4	3	7	2	4	6	13
Total	5	5	10	2	5	7	17

*) According to their own saying

Table 7.A3.3. Number of respondents who attended a vocational education or training since January 1994 *), by background characteristics.

Age	Gender						Total
	male			female			
	Education level		Total	Education level		Total	
	low	high		low	high		
< 30	1	1	2	2	1	3	5
≥ 30	4	3	7	3	3	6	13
Total	5	4	9	5	4	9	18

*) According to their own saying

Chapter 8

Case Study 3: Testing Questions from the Continuous Survey on Living Conditions

Computer-Assisted Qualitative Interviewing: Testing and Quality Assessment of CAPI and CATI Questionnaires in the Field

Ger Snijkers, Edith de Leeuw, Doortje Hoezen, and Isje Kuijpers

Summary: *Within a major redesign program of several socio-cultural surveys (POLS) at Statistics Netherlands, in 1996 an extensive test program has been carried out. This study was carried out by order of the former Department of Socio-Cultural Household Surveys. In this test program all steps of the 5-step (pre-) test model of data collection development have been followed. In this chapter I will concentrate on phase 3 of this test program: the qualitative operational field test. The newly designed, complex computerised questionnaire had to be conducted with a mixed-mode scenario (CAPI-CATI or vice versa), in order to reduce non-response. The aim of the field test was to choose one of both mixed-mode scenarios. This choice was based on response rates, and interviewer and respondent preferences, investigated by using debriefing questions among others. A second goal of the field test was to assess the quality of some survey questions. This was done by using cognitive interviewing methods (reaction coding and probing) in the field. These methods, as well as the debriefing questions, were integrated in the computerised questionnaire, resulting in a Computer-Assisted Qualitative Interviewing tool (CAQI). We feel that this study could not have been carried out without CAQI as a quality assessment tool.*

In chapters 6 and 7 we have seen examples of pre-test studies in the laboratory using CAQI (in step 2 of the 5-step (pre-)test model for data collection development). The study presented in this chapter is an example of a qualitative operational field study (step 3), using CAQI and cognitive laboratory methods in the field. This chapter is a reprint of Snijkers et al. (1999), which is an extended version of a report on this field test (Snijkers, 1996). A report on the entire test program, which included the field test report, had been presented to the client by De Heer (1996). The client judged this report as a "very useful, complete and very detailed documentation of the test program" (Everaers, 1996).

Keywords: *Cognitive Interviewing in the Field, CAPI, CATI, CAQI*

- First version of paper presented at the International Conference on Computer-Assisted Survey Information Collection (InterCASIC'96), December 11-14, 1996, San Antonio, Texas.
- Final version presented at the Third ASC (Association for Survey Computing) International Conference ('… Leading Survey and Statistical Computing into the new Millennium'), September 22-24, 1999, The University of Edinburgh, Edinburgh, Scotland, UK.
- Paper published in: Banks et al. (eds.), 1999, '… Leading Survey and Statistical Computing into the new Millennium'. The Conference Proceedings, pp. 231-258. (Association for Survey Computing, Chesham, Bucks, UK.)
- This research project has been carried out conjointly with Hans Akkerboom, Dirkjan Beukenhorst, Piet Boersma, Wim Burgers, Francine Dehue, Rob Giessen, Wim de Heer, Frans Kerssemakers and Paul van Litsenburg.

8.1. Introduction

The Questionnaire Laboratory at Statistics Netherlands started its work in 1992. This Laboratory offers facilities and methods for designing and pre-testing questionnaires. A pre-test program typically consists of a small number of qualitative interviews (say up to 50). In these interviews respondents are requested to provide information on the question-and-answer process by using cognitive interviewing techniques like thinking-aloud, paraphrasing, meaning-oriented probing, vignettes, etc.

At Statistics Netherlands, almost all household surveys are carried out by means of Computer-Assisted Personal or Telephone Interviewing (CAPI or CATI) using Blaise. Within this setting, we designed a computer-assisted method of pre-testing computerised questionnaires in a cognitive laboratory. This computer-assisted method for pre-testing questionnaires is meant to provide *Qualitative information* about how questions are processed and how answers are obtained. The CAI interview becomes a *CAQI interview*. By adding cognitive interviewing techniques to a standardised CAI interview, the original Blaise questionnaire is extended with a 'CAQI protocol': the script for applying cognitive interviewing techniques during or after the ordinary interview.

CAQI is applied in many pre-test programs. (For a more detailed description of CAQI in a laboratory setting, we refer to Snijkers, 1997.) However, in many cases it is impracticable and/or unrealistic to assess quality in a laboratory setting, e.g. when the sampling design or fieldwork procedures should be tested. Also the (redesigned) questionnaire may be tested on a larger scale in a realistic field setting, to find out whether questions are appropriate in the field to a larger number of respondents, or whether there any problems that did not show up in a previous laboratory pre-test. CAQI can be applied under the same conditions as CAI, the only difference being that CAQI focuses on qualitative information about the survey questions.

So, in addition to laboratory research, qualitative interviewing may be worthwhile in other settings on a larger scale. Gradually, CAQI evolved into a quality assessment tool that allows the cognitive laboratory to be 'moved into the field', both for household and business surveys. In this way not only the questionnaire but also all aspects of the survey design can be tested.

At Statistics Netherlands, in 1996 a pilot study has been carried out as a first attempt to test the design of a household survey (the Continuous Survey on Living Conditions: in Dutch abbreviated as POLS) using CAQI in a qualitative operational field test. In this chapter special attention will be given to the CAQI design of this study. In section 8.2 the design of this POLS test will be described, including a discussion on the application of CAQI. The results of this study are dealt with in section 8.3, including a discussion on the results. Section 8.4 concludes this chapter.

8.2. The POLS qualitative operational field test

8.2.1. Design of the POLS test program

POLS is an integrated system of social surveys, designed in 1995, tested in 1996 and implemented in the field in 1997. The incentives to compose a new integrated system were: to meet new demands concerning the analysis of social relations using a comprehensive data set, to develop a method for a major redesign of individual surveys in order to compose such a comprehensive data set, and to reduce non-response (Bakker & Winkels, 1998a).

The POLS questionnaire is composed of several individual questionnaires, which were all computerised. This new questionnaire is basically composed of two modules: the base module and the specific module. The base module consists of questions on background variables and screening questions on several topics like household composition, education level, time use, housing, health status, social participation, cultural participation, political participation, perceived safety, and income. This module lasts about 15 minutes. In order to reduce non-response this module is conducted face-to-face (CAPI) as well as by telephone (CATI) in a mixed-mode design: first a CAPI interview will be conducted; if this results in no-contact or refusal, a CATI interview will be conducted. The specific module is conducted by CAPI only, and lasts about 30 minutes. In case the base module is conducted with CAPI the specific module will follow this module immediately, otherwise the respondent has to be contacted again. In the specific module topics like health, crime, general well being, housing or time use will be addressed in more detail.

To get a well-designed survey an extensive test program was put forward, consisting of 5 phases. This test program was both aimed at testing the questions in the base module questionnaire and testing the data collection design (mixed mode). The test program gradually evolved from small-scale laboratory testing to a large-scale pilot study in which the focus shifted from exploratory qualitative results, to confirmatory quantitative ones. Correspondingly, the focus shifted from the questionnaire to the data-collection process.

The first phase of the test program was a definition study, in which the base module was developed and reviewed by a few experts. This review resulted in several issues to be investigated.

Phase two was a small-scale qualitative laboratory test of the base module in the questionnaire laboratory including an expert review, 2 focus groups and 46 CAQI interviews (Dehue, 1996). In this phase a number of issues, resulting from the first phase, were investigated: general respondent commitment (attitude and interest concerning the topics addressed in POLS), respondents' opinion on going from one topic to another in the base module (transitions), the order of the topics in the base module (which order is most preferred by respondents), and testing the base module for CAPI as well as for CATI. Apart from these issues also the individual questions were tested with regard

to the question-and-answer process. Tourangeau and Rasinski (1988) distinguish 4 stages respondents have to go through in this process: (1) interpretation of the question, (2) retrieval of information from memory needed for the answer, (3) combining the retrieved information into an overall answer (judgement stage), and (4) mapping of the judgement onto the listed response options (reporting stage). In the test these stages were addressed.

After completion of the qualitative laboratory test the emphasis shifted from the questions in the questionnaire to fieldwork procedures. In the *third phase* a large-scale qualitative operational field test was carried out (Snijkers, 1996). The focus of this test was on comparing two mixed-mode scenarios under realistic fieldwork conditions (a CAPI-CATI and a CATI-CAPI scenario) in order to make a well-based choice for one of both. The first scenario was the CAPI-CATI scenario. Respondents who participated in this scenario only had one interview in which both base and specific module were conducted. For initial non-respondents (after several trials), a telephone interview for the base module was planned, followed by an interview at home (using CAPI) to conduct the specific module. As for the CATI-CAPI scenario, a telephone interview was conducted to administer the base module, followed by a CAPI interview at home for the specific module. Initial non-respondents were re-contacted using CAPI for one interview, including both base and specific module. A second goal of this operational test was to investigate some (partially) unresolved problems from phase two with regard to the questions in the base module. In the following subsections we will concentrate on the design of this test, *as an example of CAQI in the field*.

Phase four was a large-scale quantitative pilot study. *Phase five* is the implementation phase (transition form test to survey). (This test program is described in more detail by De Heer, 1996; Akkerboom and Dehue, 1997; Akkerboom, Dehue and Snijkers, 1998.)

8.2.2. The POLS qualitative operational field test

In *the third phase* (the qualitative operational field test) several aspects of the survey were tested under realistic fieldwork conditions: the data collection design (the mixed-mode design, and fieldwork procedures like the introduction of the survey by the interviewer at the door or on the phone), and the quality of a small number of questions. For these purposes the base module was extended with several CAQI questions: questions referring to survey questions, some respondent-debriefing questions, and some interviewer-debriefing questions. Also two interviewer-debriefing sessions were organised after the fieldwork was completed.

In this test 688 households in three regions of the Netherlands were contacted (which resulted in 365 responding households). For this study individuals were sampled. The interviews were carried out in March-April 1996, in a very strict time schedule. For this test 21 CAPI and 13 CATI interviewers were carefully selected and trained. These interviewers were experienced interviewers

in standardised interviewing and familiar with Blaise. (However, in this study a new version of Blaise (Blaise III) was used for the first time.) Preferably, they also had to have specific characteristics like, being able to listen carefully and objectively, being curious (wanting to know why), being able to make short and clear summaries of remarks, and having a critical attitude towards the standardised questionnaires that are used in surveys.

The training was split up into two sessions. The first session was a one-day introduction of POLS, the field test and the task of the interviewers. The second session was a two-day training just before the start of the fieldwork. An important aspect of this training program was the task and attitude of CAQI interviewers, i.e. how to apply cognitive interviewing techniques, in particular carefully listening and non-directional probing. Furthermore, in the training program a lot of attention was given to the kind of information (i.e. qualitative information) that should be gathered in the CAQI interviews. This is because interviewers trained in standardised administration of closed questions have to become so flexible as to be able to include non-directional probing in their behaviour. Also attention was given to the use of Blaise III. (A detailed description of the design of the field test, the recruiting, selection and training of interviewers is presented by Snijkers, De Leeuw and Hoezen, 1997, and De Leeuw, Snijkers and Hoezen, 1997.)

As for the mixed-mode design, a choice for one of both scenarios had to be made. The criteria the choice was based on, were response rate, and respondent and interviewer preference for a scenario. With regard to the latter, debriefing questions for respondents and interviewers were added to the questionnaire, and interviewer-debriefing sessions were organised after the fieldwork was completed.

The debriefing questions for the respondent and the interviewer were on e.g. opinion on the length of the interview, opinion on the subjects covered in the interview (interest and attitude). Other debriefing questions for the interviewer were on the flow of the interview (pleasant/unpleasant), the co-operation of the respondent during the interview, and the burden of the interview task. (Some of these questions were based on debriefing questionnaires used by Campanelli, 1995, and Couper, 1995.) These questions were put at the end of the questionnaire, in order not to interrupt the interview.

The other CAQI questions were questions to investigate the quality of several survey questions that still had major problems after the prior qualitative laboratory study (phase two of the test program). Nine questions were selected for further investigation. After the laboratory study the question wording and/or the answering categories of these questions were slightly changed. The aim of the field test with regard to these questions was to find out whether the reformulated questions did not give rise to major problems in the field concerning the question-and-answer process. Since the CAQI questionnaire had to be used in face-to-face as well as in telephone interviews the number of CAQI questions had to be limited. The qualitative interviewing techniques used in this field test were coding of respondent reactions (based on Oksenberg, Cannell

& Kalton, 1991) and elaboration probing. These questions and the aims of the test are listed in table 8.1.

Table 8.1. POLS questions and CAQI questions

POLS question	Question wording	Aims of the test and expected reactions	CAQI question
TimeUse	What is your most time-consuming activity, apart from relaxation? Is that: paid work, house-keeping, study or education, volunteering, something else ?	Important routing question: Testing of answerability	Reaction coding
WantWork	Would you like to have paid work for 12 hours a week or more ? (yes, no, already found, would like to but cannot)	After phase 2 response option 'would like to but cannot' added to list. Important routing question: Testing of answerability	Reaction coding
Health	What is in general your health condition? Is it: very well, well, moderately good/bad, some-times good and sometimes bad, or bad?	Wording slightly changed after phase 2. Problems with answering categories are expected.	Reaction coding
Disabled	Are you hindered in doing daily activities caused by a long lasting illness, affection or disability in other situations, like going to work or in your spare time? (yes, no)	This was a third question in a row of approximately equally worded questions about the same issue. Irritated reactions of respondents and answering before end of the question are expected.	Reaction coding
Doctor	The next question is on contacting the family doctor. All visits to and by your doctor have to be counted, as well as telephone consultations. You don't have to count renewal of prescrip-tions. How many times did you for yourself contact the family doctor in the last 14 days, so since …?	Wording is slightly changed. Long introduction: It is expected that respondents request for repetition of the question and/or for clarification.	Reaction coding
NonSmoke	Do you never smoke? (never, occasionally)	Asked when answered 'no' to 'Do you smoke?'. Response options changed after phase 2. Superfluous question for non-smokers: It is expected that respondents get irritated.	Reaction coding
FamCont	The next question is on contacting members of your family. This includes meetings, telephone and written contacts with family members that don't live in your house. How many times do you contact members of your family? Is that: at least once a week, twice a month, once a month, less than once a month or seldom or never?	Wording and response options are changed. Long question and unclear meaning of 'contact' and 'members of the family': It is expected that respondents request for repetition of the question and/or for clarification.	Reaction coding
DwelSize	How many living, sleeping and working rooms are there in your house?	Difficult task and unclear meaning of 'room': Investigating interpretation and finding out whether other rooms have been counted as well (check validity and reliability).	Elaboration probe
HouseWrk	The next question is on housekeeping activities, like tidying up and cleaning, washing and ironing, shopping, cooking, and doing the dishes, and caring for plants and pets. How many hours a week do you yourself usually spend on such housekeeping activities?	Difficult task and unclear meaning of 'housework': Investigate interpretation and finding out whether the number of hours counted is a reliable number (check validity and reliability).	Elaboration probe

8.2.3. The POLS CAQI protocol

The starting point for CAQI is a computerised questionnaire (in our case the POLS Blaise questionnaire). To get a CAQI protocol, the cognitive interviewing protocol is *integrated* in this questionnaire. A CAQI protocol is expressed by *instruction screens* built around the (groups of) questions that are to be investigated and additional questions that are added to the questionnaire. The screens tell the interviewer:

(1) when to insert which cognitive interviewing techniques during the standardised interview (headings like "PAY ATTENTION: enter reaction code after this question", or "PAY ATTENTION: two elaboration probes"), and

(2) how to apply these techniques (e.g. "Enter code(s) for respondent reaction to question.").

This means that all instructions are in the computer. For conducting a CAQI interview the interviewer starts the interview program on the computer and every single step (both with regard to the standardised and the CAQI interviewing protocol) is put forward *automatically*.

In the POLS questionnaire the insertion of the reaction coding was very easy to do. This CAQI question was defined in the Blaise questionnaire only once as a special block. Thus, this CAQI question could be inserted in the routing paragraph wherever appropriate, simple by referring to this block. This resulted in a questionnaire in which the interviewer firstly had to pose the survey question and enter the answer, and secondly, immediately after the question had to enter the reaction codes. In figure 8.1 and 8.2 the computer screens of the survey question and the CAQI question are shown.

In figure 8.1 we see that at the top of the screen an interviewer instruction is printed (indicated with: >> INT) saying that the interviewer had to pay attention to the respondent's reaction and making the interviewer aware of what is to come. The rest of the screen is like the ordinary survey question. In figure 8.2 we see how the actual interviewing task is shown on the screen. At the top of the screen some interviewer instructions are listed. Here "OBSERVE" indicates that this question does not have to be asked to the respondent, but that the interviewer may enter the appropriate codes.

Figure 8.1. Standardised survey question (Time use) with interviewer instruction for reaction coding

```
>> INT: PAY ATTENTION, enter reaction codes after this question:
        How does respondent react? <<

What is your most time-consuming activity, apart from relaxation?
Is that: -->

(enter code)
1: Paid work,
2: Housekeeping at home,
3: study or education,
4: Volunteering or working with benefit, or
5: Something else?
```

Figure 8.2. CAQI screen for reaction coding

```
OBSERVE: REACTION CODE(S)
         Enter code(s) for respondent reaction to question.
         More than one code can be entered.
         In case no code is applicable, press <ENTER>.


Respondent:
1. asks for repetition of the question
2. asks for clarification of the question
3. interrupts question reading
4. expresses doubt between response categories
5. gives answer that doesn't fit in the list of categories
6. changes answer
7. gets irritated by question
```

The number of codes is limited, since the coding had to be done during the interview. It would have been impossible for interviewers to remember a long list of codes during the interview. Furthermore, it was expected that a long list would disturb the interview, especially the telephone interview. These codes were chosen because they cover the whole question-and-answer process (Tourangeau & Rasinski, 1988; Clark & Schober, 1992; Fowler & Cannell, 1996):

- The first three codes deal with the interpretation stage. The first two codes indicate that the question may be hard to understand, due to unclear or undefined terms, long question, complex syntax, poor question order (influencing the interpretation), unclear response task, or respondents cannot recall response categories. The question may not be valid.
  The third code may indicate that apparently there is no interpretation problem, the question may be too long and redundant, or the respondent doesn't realise that response categories will be given.
- The following two codes have to do with the list of answering categories (the judgement and reporting stage): it may be hard for the respondent to make up his mind and choose for only one option (the categories may not be clearly mutually exclusive), or response alternatives may be missing (the list may not be exhaustive).
- The sixth code indicates difficulties with regard to retrieval and judgement of information: it may be hard for respondents to remember the correct information; the answer may not be reliable.
- The last code expresses a general feeling of irritation by the respondent towards the question.

With the questions on dwelling size ('DwelSize') and housekeeping activities ('HouseWrk') elaboration probes were posed to the respondents. These CAQI questions were defined as ordinary survey questions in the Blaise questionnaire. As for 'DwelSize' the elaboration probe was asked immediately after the survey question itself. In an additional CAQI question it was possible to change the answer to the survey question. The questions as they appeared on the computer screen are shown in figures 8.3, 8.4 and 8.5.

Figure 8.3. Standardised survey question ('DwelSize') with interviewer instruction for elaboration probe

```
>> INT: PAY ATTENTION, two elaboration probes will follow. <<

How many living, sleeping, and working rooms are there in your house?

(enter number between 1 and 97)
```

Figure 8.4. Elaboration probe with number of rooms

```
Did you include the kitchen, the bathroom and/or the cellar in the number
of rooms?

>> INT: You are not allowed to change the answer to the last question.
        In case the respondent wants to change the number of rooms,
        you can enter that number right after this. <<

(enter code)
1: Yes
2: No
```

Figure 8.5. Observe-question to change answer

```
OBSERVE: In case the respondent wants to change the number of rooms,
         you may enter this number here.
         The number given before was: ^Rooms.
         In case the respondents HIM/HERSELF does not indicate any changes,
         press <ENTER>.

(enter number between 1 and 97)
```

The elaboration probes with 'HouseWrk' asked for the time spend on every individual house keeping activity listed in the survey question (see table 8.1). These probes (see figure 8.6) were asked at the end of the base module, to check the reliability of the former answer. In between the original question and the elaboration probes, there were about 60 questions on housing, health conditions and leisure time activities. At the end of the probes it was also possible to change the former answer. An example of such a probe is presented in figure 8.6 (on the screen these questions looked like the ones shown in figures 8.3 - 8.5).

Figure 8.6a. Elaboration probe with housework (number of hours spend on tidying up)

```
>> INT: You are not allowed to change the answer to the question on
        housekeeping activities.
        In case the respondent wants to change the number of rooms,
        you can enter that number right after these probes.

Now I would like to come back to housekeeping activities.
How many hours a day are you usually busy with tidying up and cleaning?

>> INT: enter number of hours. <<

(enter number between 0 and 24)
```

Figure 8.6b. Elaboration probe with housework (number of minutes spend on tidying up)

```
>> INT: enter number of minutes. <<
(enter number between 0 and 60)
```

8.2.4. Experiences with CAQI in the POLS qualitative operational field test

8.2.4.1. Debriefing interviewers

In two debriefing sessions the experiences of the interviewers with the CAQI questionnaire were evaluated. One session was with the CAPI interviewers, while the other was with the CATI interviewers. In both cases reactions were the same.

As for the reaction coding, the interviewers did not think that this interviewing task interrupted the interview or would make it last much longer. It was surprising to hear this also from the CATI interviewers, since it was expected that this task would disturb the telephone interview. In general, the phone interview did not take too long.

However, the interviewers thought that there was too much text on the screen for reaction coding: the labels for the codes were too long. At the beginning, they were overwhelmed by this screen: “What is all this?” they uttered. At that point, they also entered some codes quickly, not to let the respondent wait too long: “You don’t have the time too read it all during an interview, since you should not have the respondent wait too long!” Here, we see that interviewers try to make the interview as pleasant as possible for the respondent. But, we have the impression that in the first interviews they pressed the enter button fairly easy to go on, since no code was applicable (see figure 8.2). Therefore, it would have been better when also a code for ‘no reactions’ would have been present in the code list. Now, we don’t know whether a missing code is non-response or ‘no reaction’. (See further subsection 8.2.4.2.)

After a few interviews, it was no trouble for them to enter a reaction code immediately after the survey question was answered. They knew when reaction coding was coming up and they knew the codes by heart. Some of them had studied the codes carefully in advance and had written the codes on a piece of paper. So, they knew when this question appeared on the screen and what reactions to look for.

In general the interviewers remarked that the codes listed were sufficient. However, they thought that the codes were not applicable to the seven survey questions selected. With some other survey questions they perceived more respondent reactions that were in accordance with the codes. So, they thought it was a pity that they could not open the reaction-coding question with these survey questions. However, some of them entered the reactions in a comment window. During the

debriefing sessions they discussed their experiences with these other questions. Here, the CAPI and CATI interviewers mentioned the same questions.

These reactions from the interviewers gave the impression that a side effect of the training as CAQI interviewer was that they are more alert on problems with questions. And besides, their information on these questions is more precise: they do not only say that a particular question is not clear to respondents, but they also try to give a precise indication of what is not clear.

As for the elaboration questions, the elaboration question with 'Rooms' did not give rise to any serious problems. However, the interviewers did not understand the intention of this CAQI question, which was to find out whether the interpretation of the term 'rooms' is in accordance with the rooms that have to be counted. The elaboration probes with regard to housework were annoying to both the interviewer and the respondent. To ask for these activities on a daily basis is too detailed.

From the interviewer reactions it appeared that respondents do not interpret these elaboration questions as CAQI questions. To them, these questions are like ordinary survey questions. This means that for these questions the same question-and-answer process holds, and the same measurement errors might occur. So, they should be tested in the same way as survey questions. As for the possibility to change answers, according to the interviewers respondents do not recall the answer they had given before. That is, only a few respondents changed their former answer. As the interviewers remarked: "They do not know that the answer can be changed, so they don't."

A general side-effect of this project was that the interviewers had the feeling that they were really important in the data collection process and that their view was accounted for. They also had the impression that the respondents were enthusiastic about this study.

8.2.4.2. Learning effect in the reaction codes

As mentioned above, in the debriefing sessions we got the impression that a learning effect is present in the data on reaction coding: in the beginning of the fieldwork fewer codes were entered than afterwards. To check for such a learning effect, spearman rank-order correlations r_s (Siegel & Catellan, 1988) have been calculated for the coding (number of codes per interview) and a rank number of the day of the interview, for both modes and for the data set as a whole. We find that the correlations are very small and that none of them are significantly different from zero (at a .05 level for a two-tailed test):

- CAPI: $r_s = -.09$ ($p = .116$; $n = 290$)
- CATI: $r_s = .06$ ($p = .394$; $n = 101$; for 33 CATI interviews the interviewing date was missing (see table 8.3))
- CAPI + CATI: $r_s = -.09$ ($p = .332$; $n = 391$)

So, we may conclude that there is no learning effect.

8.2.5. Other experiences with CAQI in qualitative operational field tests: 'Vacancies to be Difficult to Fulfil'

CAQI has also been applied in another study at Statistics Netherlands. In 1996 a test on 'Vacancies to be Difficult to Fulfil' (VDF) has been carried out (like with the POLS questionnaire) both in the laboratory and in the field (Kuijpers & Akkerboom, 1997; Lammers et al., 1997).

The question on VDF is part of a quarterly survey for government organisations to measure vacancies: The Quarterly Survey on Employment and Wages. On a form, these organisations have to list their vacancies, according to a specific definition of what is to be considered a 'vacancy'. Once a year a few additional questions are asked about the nature of these vacancies. One of these additional questions is on VDF. In the questionnaire the concept of VDF is not further explained, which may result in wide interpretations of the concept. In order to investigate the problems with this question and to improve measurement of vacancies with government organisations, a test program was carried out. This was done in the laboratory by a series of in-depth interviews and focus groups, and in the field with 150 business interviewers. These interviewers used a Blaise questionnaire that was developed on the basis of the laboratory test results. This Blaise questionnaire consisted of vignettes on vacancies and VDF's, open-ended comprehension probes to investigate how certain concepts are interpreted, and debriefing questions with confidence rating of answers. Half of the CAQI interviews were telephone interviews, the other half face-to-face.

Like in the POLS test, after the interviews were conducted, debriefing sessions with the interviewers were organised. Again, the interviewers were asked to tell about their experiences with the CAQI questionnaire. The main aspects they referred to were the use of the computer and the design of the questionnaire. Some aspects concerning interviewing techniques were also mentioned. In general, the experiences reported in this study are comparable to the ones reported by the POLS interviewers (as discussed in subsection 8.2.4).

As for the use of the computer, the interviewers indicated that it is important to know how to handle the computer. They were used to work with paper forms, and now they had to use a computer. Sometimes the continuation of the interview was interrupted because the interviewers did not know how to handle the computer or the Blaise questionnaire. This was for instance the case with open questions. It was difficult to enter summaries, since most of the interviewers were not used to typing. Some of them solved this problem by writing the summaries on paper, and entering these into the computer afterwards. They also thought that it was difficult to listen to the respondent and enter a summary at the same time. This indicated that further training of the interviewers in this respect was necessary.

Some interviewers also had the impression that the computer is impersonal and detached. That is why some interviewers decided to have the respondent sit next to them and watch the screen.

However, as other interviewers remarked, then the interview will be disturbed by the information on the screen that is not meant for the respondent.

Other interviewers said that the computer gave a professional and modern impression to the respondent. Since the data were entered into the computer immediately, this gave the idea that the data are really used. Furthermore, the respondents were very enthusiastic about the interviews as a whole. They were very satisfied that their views were taken account of. In general, they did not think that the interview was annoying or that it took too much of their time.

With regard to the design of the questionnaire, the interviewers remarked that the space they had for a summary was too limited. This was also a reason why they wrote summaries on paper and entered these in the computer afterwards. However, space was limited on purpose, to force interviewers to enter the most essential aspects the respondent mentioned. This indicates that they had difficulties with making proper summaries during the interview.

As for the presentation of the probes on the screen, some interviewers used all of them during the interview. During the debriefing they remarked that some respondents got irritated with all their probing. These interviewers did not understand the intention of the probes on the screen, i.e. to use them in a selective way whenever necessary. Others thought that the listed probes were not appropriate, and appropriate ones were missing. They did not use any of them nor any other probe. This gave the impression that if the probes had not been listed on the screen, no one would have probed at all, so that only very limited information would have been obtained form the CAQI interviews. So, probes have to be listed on the screen, however it is important to give convincing but strict instructions to the interviewers.

Another remark concerning the probes was that there was a lot of text on the screen. At first, the interviewers did not know where to look on the screen, what to read aloud, and what should only be read by themselves. The conclusion again is that it is very important to train the interviewers in the use of the questionnaire. They have to get acquainted with the questionnaire and the layout of the screen. It would have been helpful if colours could have been used, but the laptop computers used at that time did not offer such an option.

8.2.6. Discussion and conclusions: Evaluation of CAQI

Above we have described two studies in which CAQI has been applied in the field. The CAQI protocols have been develop on the basis of earlier small-scale qualitative research. Here, we will summarise our experiences with CAQI.

While planning and executing a CAQI field test like in the POLS case, a good collaboration between the researchers and the staff from the fieldwork department is necessary. At Statistics

Netherlands the Fieldwork Department is responsible for recruiting and training interviewers, the programming of the Blaise questionnaire and the organisation (including sampling) and execution of interviews according to standard procedures. The Department of Data Collection Methodology was responsible for the design of the test program. In the POLS qualitative operational field test both researchers and fieldwork staff combined their expertise to design and carry out this test.

During the preparation stage attention should be given to the design of the questionnaire: the CAQI questionnaire should be carefully developed like all CAI questionnaires (Snijkers, 1992, chapter 2), and it should be tested in test runs to detect programming errors and in usability tests with interviewers (Stratton & Hardy, 1996). This may be quite time-consuming. On the other hand, in the CAQI questionnaire all features of CAI questionnaires are included: like quickly changing the questionnaire e.g. in case of a changed design.

Now, to summarise the preparational aspects:

- Good co-operation between research and fieldwork staff is necessary to accomplish:
  - recruiting and training of interviewers,
  - organisation and execution of interviews according to standard procedures,
  - availability of CAI questionnaire to built in the CAQI protocol.
- Preparation of CAQI questionnaire:
  - all stages of the CAQI protocol should be carefully prepared and tested,
  - time-consuming,
  - the questionnaire can be changed quickly.

After the field interviews had been conducted, the CAQI method had been evaluated using debriefing sessions with interviewers, in order to improve the application of the CAQI. The comments of the interviewers may be classified into three groups: comments having to do with the use of the computer, the interviewer and the respondent.

To carry out pre-test studies like the POLS and DVF studies, the use of the computer is important. The POLS qualitative field test (with this number of respondents and this complex ‘mixed-mode’ design) would have been practically impossible if the probes were not included in the CAI questionnaire. Due to a complex routing structure, it was impossible to get a paper version of the questionnaire. Another option would have been that a paper version of the cognitive protocol had to be used along with the computer. In that case, the interviewer had to pay attention to both instruments and to the respondent as well. In the CAQI interviews however, the survey and the CAQI questions are both in one and the same computerised questionnaire. Thus, the interviewer is relieved of the burden of following the correct routings, and consequently can give more attention to the respondent. The computer helps the interviewer to control the flow of the interview, both for the survey questions (like with CAI as described by Snijkers, 1992, chapter 2; De Leeuw, Hox and Snijkers, 1998, chapter 3) and for the CAQI protocol.

Another argument for using the computer is that with CAQI instructions for the interviewer appear on the screen. This prompts the interviewer to apply the appropriate cognitive interviewing

techniques at the right time. The remarks of the interviewers in the debriefing sessions make clear that the instructions on the screen make the interviewers use the probes. Analysing learning effects for the reaction codes revealed that there are no such effects, so the data seem to be reliable.

The cognitive interviewing techniques that have been applied are reaction coding, closed and open-ended probing, vignettes and debriefing questions. Apart from reaction coding, these techniques are also described by DeMaio and Rothgeb (1996) as techniques to be used in the field. In accordance with our experience they point out that not too many problems should be addressed in the field, in order to minimise the additional burden for the respondent and the interviewer. They also based their field questionnaires on former qualitative research.

So to summarise, CAQI in the field is characterised by:

- Greater scope of qualitative research:
  - qualitative research to be used in the field, using reaction coding, probing, vignettes and debriefing questions,
  - more data: larger number of respondents,
- Integration of the cognitive interviewing protocol in the standardised CAI questionnaire, thus:
  - conducting cognitive interviews in a structured way,
  - focussing on respondent's reactions.
- Reflection of standard fieldwork conditions, by using computer-assisted interviewing and its features, like:
  - complex skipping patterns,
  - data entry during the interview.
- Control of the flow of the interview by the computer:
  - Resulting in appropriate probing: the appropriate cognitive interviewing techniques appropriately applied at the right time (what technique when and how).
  - Thus, gathering reliable qualitative information with regard to the aim of the test program.
  - But, to achieve this:
    - every single action should be on a separate screen,
    - there should not be too much text on the screens.

As for the appropriate application of cognitive interviewing techniques, it is not enough to show on the screen what should be applied when and how. In order to make sure that CAQI interviews are conducted in the right way, it is important that interviewers are well trained in applying (cognitive) interviewing techniques and are instructed in why a probe or question is added to the questionnaire. Also the interviewers stress the importance of being properly trained in handling the computer (like typing, use of the keyboard, basics on the operating system) and the computerised questionnaire (like data entry, changing data, function keys, paging). The importance of interviewer training, including these issues, is also stressed by Wojcik and Hunt (1998), De Leeuw, Hox and Snijkers (1998, chapter 3) and Morton-Williams (1993).

We find that when interviewers are involved in the process of designing a survey at an early stage, and are given attention in the form of instructions and training, they get the feeling that they are an essential part of the process, they will be well informed, and thus they will perform a better

job. All in all, the interviewers were very enthusiastic about their task, which was more than gathering survey data only.

As for the interviewers, we may summarise:

- During the interview interviewers have to pay attention to:
  - the survey questions,
  - the cognitive protocol,
  - the computer,
  - the respondent.
- To do a good job they should be well trained in:
  - properly applying (cognitive) interviewing techniques,
  - the goals (the why) of the test program,
  - handling the computer and the computerised questionnaire.

In both studies the interviewers had the impression that the respondents did not think that the CAQI questions were annoying. Respondents did not think of the CAQI questions as test questions, but as ordinary survey questions. They expressed a feeling that their opinions could be helpful to Statistics Netherlands. This attitude was of great importance to the atmosphere during the interviews, which was one of great co-operation. Furthermore, the use of the computer gave the interview a professional appeal. (This is a general attitude of the respondent with regard to computer-assisted interviewing; as described by De Leeuw, Hox and Snijkers, 1998, chapter 3).

With regard to the respondent we may summarise that:

- The use of the computer in qualitative interviews gives the interviews a professional status, in which spontaneous reactions still are possible.
- The respondent had the feeling that his opinion is listened to.

8.3. Results of the POLS qualitative operational field test

As stated in section 8.2.2 the aims of the POLS operational test were to make a choice for a mixed-mode design and to assess the quality of some survey questions in the base module. In this section the results will be described.

8.3.1. Choosing one mixed-mode scenario

The choice for one mixed-mode scenario is based on response rates, interviewer and respondent preferences. In summary, response rates and interviewers' opinions argued in favour of the CAPI-CATI scenario. The response rates for this scenario were, according to Dutch standards, fairly decent. The respondents did not show a clear preference for one scenario. Thus, for the survey the CAPI-CATI scenario has been chosen.

8.3.1.1. Response rate

For the CAPI-CATI scenario the net sample size was 338; for the CATI-CAPI scenario 298, excluding 52 addresses without telephone or with an unlisted number (which were contacted using CAPI). All together, the base module and the specific module were administered in one interview with 290 respondents (using CAPI); the base module on its own was administered with 134 respondents on the phone (CATI), which resulted in 75 additional CAPI interviews for the specific module. In table 8.2 the various numbers are shown.

Table 8.2. Response on mixed-mode scenarios

Mixed-mode scenario	Response numbers and rates				Net sample size
	Base module			Base + specific module	
	CAPI	CATI	Total		
	n (%)	n (%)	n (%)	n (%)	n (%)
CAPI-CATI	220 (65)	14 (4)	234 (69)	222 (66)	338 (100)
CATI-CAPI	40 (13)	120 (40)	160 (54)	113 (38)	298 (100)
CATI-CAPI (unlisted)	30 (58)	- (-)	30 (58)	30 (58)	52 (100)
Total	290 (42)	134 (19)	424 (62)	365 (53)	688 (100)

In terms of response rates, the CAPI-CATI scenario performed better than CATI-CAPI. To get a complete set of data (base and specific module), 66% response was reached with the CAPI-CATI scenario, and 38% with CATI-CAPI. Also for the base module only, the CAPI-CATI scenario resulted in higher response rates: 69% compared to 54% in the CATI-CAPI scenario. All together, the figures in table 8.2 indicate that:

1. gathering all data in one interview (CAPI: base + specific module) resulted in higher total response rates than gathering data by two interviews (CATI: base module, CAPI: specific module), and
2. response rates were higher when respondents were initially contacted at the door (65%) than when initially contacted by phone (40%).

However, in interpreting these figures it is important to keep in mind that this was a first test in the field with regard to POLS. The test brought to light a lot of operational problems, which especially affected the response rates of the CATI-CAPI scenario adversely. These problems were:

1. a very tight time schedule, making it impossible to re-contact all initial non-respondents for the base module and respondents for the specific module,
2. a lacking of experience with mixed-mode designs among interviewers (concerning the attunement between CATI and CAPI interviewers) and among automation staff (concerning data transmission and processing).

8.3.1.2. Respondent and interviewer preference

To find out what scenario respondents and interviewers would prefer most, at the end of each interview a small number of debriefing questions were asked to the respondent and some debriefing questions had to be filled in by the interviewer. These questions concerned the length of

the interview, the interest of the respondent in the questionnaire, the understanding of the questions, the respondent's co-operation during the interview, the interviewer's opinion on the flow of the interview and the burden of the interview task. The respondents who participated in the CATI-CAPI scenario also had to give their preference with regard to the interviewing mode (at home or by telephone).

The results of these debriefing questions are presented in table 8.3. While looking at this table it is important to keep in mind that the columns represent different combinations of modules and interviewing modes. However, regarding each column as a different scenario, the debriefing questions do not reveal a clear preference for one of both scenarios, neither for the respondent nor for the interviewer.

Table 8.3. Respondent and interviewer opinions on the mixed-mode scenarios

Aspect of interview	Combination of module and interviewing mode		
	Base + specific module (CAPI)	Base module (CATI)	Specific module (CAPI)
Total number of interviews	290	134	75
Average length (min.)	45	15	30
	%	%	%
Respondent on interview:			
• Length of interview: good or could be longer	86	84	93
• Interest in questionnaire: interesting or easy	79	86	76
Interviewer on respondent:			
• Length of interview for resp.: good or could be longer	78	83	95
• Interest in questionnaire: moderate or above	94	89	97
• Understanding the questions: no problem	76	78	83
• Co-operation during interview: no problem	82	80	84
Interviewer on the interview:			
• Flow of the interview: pleasant	93	88	93
• Tiring for interviewer: not tiring	72	87	81

The various questionnaires do not seem to be too long, although the interviewers feel that conducting base (87% not tiring) and specific module (81%) separately is less tiring than conducting both modules in one interview (72% not tiring). Also, the interviewers feel that a combined interview is more tiring to the respondents (78%: length is good or could be longer) than the respondents feel themselves (86%: length is good or could be longer). As for these and the other aspects in table 8.3, the percentages do not differ greatly between the various scenarios neither for interviewers nor for respondents. Hence they are not very conclusive as to the choice of a scenario. This corresponds with the findings of the qualitative laboratory test (phase 2), which showed no differences between the mode used for conducting the base module (Akkerboom, Dehue & Snijkers, 1998).

As for the respondents, these results are in accordance with remarks of the interviewers in the debriefing sessions. The interviewers remarked that respondents do not know what other scenario's there are: "They do not know better." So, they are not able to a make a choice However, the 75

respondents who did the CATI-CAPI scenario had participated in both modes. They were asked to report their preference. About 40% reported that they liked the interview the way it was (in two sessions), also about 40% would prefer one interview at home, and about 20% said that they would like one telephone interview. Again, these figures are not very conclusive as to what scenario to choose.

Despite these results, in the interviewer debriefing sessions the interviewers expressed a clear preference for the CAPI-CATI scenario. They came up with three reasons. First of all, the possibilities of refusal conversion with non-respondents were better in this scenario. In the field the same interviewer would visit non-respondents again, while in the CATI procedure non-respondents could seldom be re-contacted by the same interviewer. Moreover, in the field there was less time pressure during the introduction at the door than on the phone. Therefore, by phone it was much more difficult to gain non-respondent's confidence and co-operation. Secondly, in the CATI-CAPI scenario it was always necessary to re-contact the respondents for the specific module. This resulted in additional non-response, even if respondents on the phone had agreed to participate in a follow-up CAPI interview. Thirdly, according to the interviewers the process of switching from one mode to the other seemed to work out better in the CAPI-CATI scenario.

8.3.2. Quality assessment of the questions in the POLS base module

As for the questions in the base module, nine questions were selected for further investigation of (partially) unresolved problems in phase 2, as mentioned before. To investigate these questions respondent reactions were coded or an elaboration probe was asked. The reaction coding confirmed our expectations as formulated in table 8.1. Almost no differences were found between CAPI and CATI, which corroborates earlier results of the laboratory test (phase 2 of the test program, as discussed in subsection 8.2.1). As for the other two questions the elaboration probes revealed that, although the questions have been reformulated after the laboratory test, some problems with regard to the question-and-answer process still remained.

8.3.2.1. Reaction coding

In table 8.4 the percentages of respondents that showed any problem with the seven questions (for which the reactions were coded) are presented for the two modes. The questions are presented in the order they were asked. What strikes immediately is that for the questions the percentages do not differ greatly between CAPI and CATI, which is in accordance with the results in subsection 8.3.1. However, the registered reactions may be different.

Most reactions were registered for the question *'Time use'* (32%). This question was not changed after phase 2. In table 8.5 the number of registered codes are presented. In total, about half (42%) of the respondents who showed any problems with this question asked for repetition of the question or for some clarification. About one-third (35%) showed any difficulty in answering the

Table 8.4. Reaction coding for seven questions

Question (*)	Number and percentage of respondents that showed any problem for interviewing mode			Number of respondents question was asked to for interviewing mode		
	CAPI	CATI	Total	CAPI	CATI	Total
	n (%)	n (%)	n (%)	n	n	n
Time use	104 (36)	32 (24)	136 (32)	290	134	424
Want paid work	24 (24)	11 (24)	35 (24)	99	45	144
Health	68 (23)	29 (22)	97 (23)	290	134	424
Disabled	58 (20)	25 (19)	83 (20)	290	134	424
Contacted family doctor	45 (16)	13 (10)	58 (14)	290	134	424
Non-smoking	21 (11)	9 (10)	30 (11)	193	90	283
Contacted family members	53 (18)	35 (26)	88 (21)	290	134	424

(*) See table 8.1 for the question wording.

question: they did not know what response option to choose or could not find the appropriate option. The rest answered already while the question was still being read. To them the answer was clear, even before the response options could have been read.

These results show that, although this is a short question and the answering categories were read aloud, it may be hard for respondents to understand the meaning of the question or find an answer that fits the listed response alternatives. But, from these data we have no indication what may cause the problems. However, the results from the laboratory test may give an explanation (Dehue, 1996). These results indicate that the term *'most'* may be overlooked, so that respondents come up with more than one activity. When they have to make a choice for one activity it may be difficult to decide what they do most. Also, we could argue that the sub-sentence *'apart from relaxation'* may be overlooked, so that respondents may answer that they are usually doing nothing in particular. However, since this is not a qualified answer, respondents have to come up with another activity. They could also answer *'something else'*, but our experience from other testing programs is that respondents like to give a qualified answer in stead of *'something else'*. So, for this question we may conclude that the answerability may be difficult with regard to the interpretation, judgement and reporting stage (Tourangeau & Rasinski, 1988).

Table 8.5. Reaction coding for 'Time use'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	24 (23)	3 (9)	27 (20)
Asks for clarification of question	23 (22)	7 (22)	30 (22)
Interrupts question reading	20 (19)	9 (28)	29 (21)
Expresses doubt between options	25 (24)	6 (19)	31 (23)
Answer doesn't fit in list of options	9 (9)	7 (22)	16 (12)
Changes answer	3 (3)	0 (0)	3 (2)
Gets irritated	0 (0)	0 (0)	0 (0)
Total	104 (100)	32 (100)	136 (100)

$\chi^2(5) = 8.08$ ($p = .152$)

The second question reactions were coded for is *'Want paid work'*. As result of phase 2 a response option seemed to be missing: 'would like to, but cannot'. (Respondents who were disabled or retired answered 'yes', which resulted in inappropriate follow-up questions (Dehue, 1996).) For the field test this option was added to the list to make it exhaustive. Table 8.6 shows that the list of response options now seems appropriate, since the code 'answer doesn't fit' was registered only once.

But, still some other problems exist with this question. Altogether, with 24% of the respondents (this question was asked to) a reaction was coded. These reactions have to do with the interpretation of the question (34%) and hesitation between categories (34%). Also 23% answered before the question was read aloud fully. Again, it is difficult to argue what these problems might be. The problem could be the interpretation of *'paid work'*, e.g. paid apprenticeship or paid training (Snijkers, 1995). The problem could also be *'12 hours or more'*, in case respondents sometimes work for more than 12 hours and sometimes less (Snijkers, 1995). The registered codes indicate problems with the interpretation, judgement and reporting stage. These problems could be solved (partially) by reading the response alternatives aloud, thus setting the reference frame (Clark & Schober, 1992).

Table 8.6. Reaction coding for 'Want paid work'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	3 (13)	1 (9)	4 (11)
Asks for clarification of question	4 (17)	4 (36)	8 (23)
Interrupts question reading	6 (25)	2 (18)	8 (23)
Expresses doubt between options	8 (33)	4 (36)	12 (34)
Answer doesn't fit in list of options	1 (4)	0 (0)	1 (3)
Changes answer	0 (0)	0 (0)	0 (0)
Gets irritated	2 (8)	0 (0)	2 (6)
Total	24 (100)	11 (100)	35 (100)

$\chi^2(5) = 2.91$ ($p = .715$)

After the laboratory test the question on *'Health'* is slightly changed. A quarter of all respondents showed some difficulty with this question. From these, 35% hesitated between categories, as can be seen in table 8.7. For these respondents it may be hard to make up their mind on their health condition and choose for one response alternative that expresses their health condition accordingly. This may indicate that (some of) the response options are not exclusive or exhaustive, as was expected on the basis of the laboratory test (Dehue, 1996). Some respondents reported an answer that was not in the list, saying that the ones in the list did not express their health condition properly. Also some field interviewers reported such comments made by respondents. Table 8.7 also shows that 50% of the respondents interrupted the reading. Apparently they had no problems in reporting an answer to this closed question.

Table 8.7. Reaction coding for 'Health'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	3 (4)	1 (3)	4 (4)
Asks for clarification of question	3 (4)	1 (3)	4 (4)
Interrupts question reading	31 (46)	17 (59)	48 (50)
Expresses doubt between options	26 (38)	8 (28)	34 (35)
Answer doesn't fit in list of options	1 (1)	0 (0)	1 (1)
Changes answer	2 (3)	2 (7)	4 (4)
Gets irritated	2 (3)	0 (7)	2 (2)
Total	68 (100)	29 (100)	97 (100)

$\chi^2(6)= 3.50$ ($p = .744$)

A question that was reformulated after the qualitative laboratory test (phase 2) is the question on being *'Disabled'*. In the new version this is a set of three long questions (on disabilities at home, at work, or in other situations), posed immediately after each other. The wording of these three questions is very similar, differing only on the location to which it applies, indicated at the very end of each question. Reactions were registered for 20% of the respondents. The registered reactions show that more than half of these respondents did not wait till the end of the question and interrupted the interviewer reading the question, as can be seen in table 8.8. Also, another 11% is irritated by this question. Almost all of the respondents who answered early (43) or got irritated (7) are not disabled. This indicates that the question is redundant and that respondents would like to go on with the interview (as well as for CAPI as for CATI), especially when they are not al all disabled. These results confirm our expectations.

Yet, another 17% expresses doubt between categories. From comments reported by interviewers from the field we know that some respondents answered "occasionally". This indicates that the response alternatives are not exhaustive, according to the respondents.

Table 8.8. Reaction coding for 'Disabled'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	2 (3)	0 (0)	2 (2)
Asks for clarification of question	2 (3)	5 (20)	7 (8)
Interrupts question reading	31 (53)	16 (64)	47 (57)
Expresses doubt between options	11 (19)	3 (12)	14 (17)
Answer doesn't fit in list of options	2 (3)	0 (0)	2 (2)
Changes answer	2 (3)	0 (0)	2 (2)
Gets irritated	8 (14)	1 (4)	9 (11)
Total	58 (100)	25 (100)	83 (100)

$\chi^2(6)= 10.65$ ($p = .100$)

For the question on *'Contacted family doctor'* the percentages are presented in table 8.9. Of the respondents, 47% asked for clarification or repetition, and 24% hesitated about their answer. This

indicates that this question may be hard to understand (interpretation). Also it may be hard to come up with the appropriate answer (retrieval of information from memory). From the laboratory test we know that it is not clear what person is meant in the question and whether making an appointment by phone has to be counted as a telephone consultation (Dehue, 1996). In table 8.9, we also see that 34% interrupts question reading: almost all of them (17) did not contact the doctor in the last 14 days. This is also in accordance with the results of the laboratory test, in which answers were given during the reading of the (long) introduction and the question itself. This indicates that the introduction is too long, as was expected.

Table 8.9. Reaction coding for 'Contacted family doctor'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	8 (18)	3 (23)	11 (19)
Asks for clarification of question	10 (22)	6 (46)	16 (28)
Interrupts question reading	16 (36)	4 (31)	20 (34)
Expresses doubt between options	5 (11)	0 (0)	5 (9)
Answer doesn't fit in list of options	2 (4)	0 (0)	2 (3)
Changes answer	4 (9)	0 (0)	4 (7)
Gets irritated	0 (0)	0 (0)	0 (0)
Total	45 (100)	13 (100)	58 (100)

$\chi^2(5)= 5.49$ ($p = .359$)

For 11% of the respondents (30) a reaction was coded with *'Non-smoking'*. Table 8.10 shows that 60% answered early. This is in accordance with our expectations and the results of the laboratory test: for non-smokers this question is redundant and might be interpreted as a check question (Dehue, 1996). What is striking in table 8.10 is that this reaction was almost only registered in the field; on the phone this reaction was only registered once.

On the phone about half of the respondents had a problem with the response options. This might indicate that this question is asking for sensitive information (social desirability). And thus could also be interpreted by the respondents as a check question.

Table 8.10. Reaction coding for 'Non-smoking'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	0 (0)	0 (0)	0 (0)
Asks for clarification of question	1 (5)	3 (33)	4 (13)
Interrupts question reading	17 (81)	1 (11)	18 (60)
Expresses doubt between options	1 (5)	3 (33)	4 (13)
Answer doesn't fit in list of options	0 (0)	0 (0)	0 (0)
Changes answer	0 (0)	2 (22)	2 (7)
Gets irritated	2 (10)	0 (0)	2 (7)
Total	21 (100)	9 (100)	30 (100)

$\chi^2(4)= 18.36$ ($p = .001$)

The wording of the question on *'Contacted family members'* is changed after the laboratory test: the meaning of the term *'contact'* is made clear by adding examples and the term *'family member'* is explained. This made the question longer. Also the response options are changed (Dehue, 1996). But still it is expected that the question is not clear to respondents. A reaction was registered for 21% of the respondents. About half of them asked for clarification (see table 8.11; on the phone this number is higher than for CAPI). This indicates, as expected, that the meaning of the question may not be clear, especially the terms *'contact'* (what is counted as a contact) and *'member of the family'* (who is counted as a family member) may give rise to interpretation problems. This is confirmed by comments from interviewers. A quarter of the respondents interrupts question reading, indicating that the question may be too long.

Table 8.11. Reaction coding for 'Contacted family members'

Reaction of respondent	Interviewing mode		Total
	CAPI	CATI	
	n (%)	n (%)	n (%)
Asks for repetition of question	5 (9)	2 (6)	7 (8)
Asks for clarification of question	20 (38)	22 (63)	42 (48)
Interrupts question reading	19 (36)	3 (9)	22 (25)
Expresses doubt between options	6 (11)	4 (11)	10 (11)
Answer doesn't fit in list of options	0 (0)	3 (9)	3 (3)
Changes answer	3 (6)	1 (3)	4 (5)
Gets irritated	0 (0)	0 (0)	0 (0)
Total	53 (100)	35 (100)	88 (100)

$\chi^2(5)= 14.34$ ($p = .014$)

8.3.2.2. Elaboration probes

An elaboration probe, in the form of a closed check question, referred to the answer on the reformulated question on dwelling size (*'DwelSize'*): *'How may living, sleeping and working rooms are there in your house ?'* The original question was: *'How many rooms are there in this dwelling ?'* An interviewer instruction gave a precise description of rooms *not* to be counted. This clarification was allowed to be given only at request of the respondent. In the laboratory test (Dehue, 1996), this question gave rise to problems with regard to the interpretation of the word *'rooms'*. This term was too vague. Thus, the question was reformulated by naming explicitly the rooms that had to be counted.

In the field test, immediately after the question, it was checked whether respondents had counted the kitchen, the bathroom and/or the cellar by asking the question: 'Did you count the kitchen, bathroom and/or the cellar?' About 20% of the 424 respondents mentioned that they had indeed included rooms other than the ones mentioned in the question. The check question also evoked a spontaneous revision of the original answer (9% of 424), even though some of these respondents (6% of 424) did not include other rooms than those mentioned in the question. These results indicate that (in accordance with the results of the laboratory test; Dehue, 1996) a uniform

interpretation of the concepts referring to the intended rooms appears hard to obtain: a kitchen or a study may be viewed as a working room, a large cellar as a living room, a kitchen annex living room as a living room, and a large living room that was originally two rooms may be counted both as one or two rooms. This is confirmed by interviewer comments from the field.

In general it may be difficult for respondents to come up with the right number of rooms: to count the rooms that match the restrictions in the question, without forgetting any. But we feel that the new question is a better one than the original: the question tells specifically what rooms are to be counted (although it may not be clear whether a room is in accordance with the definition), in stead of mentioning the rooms not to be counted in a instruction (which means that respondents who do not get to hear the instruction make up their own definition).

With regard to the reformulated question *'Time used for housekeeping' ('HouseWrk'*), the elaboration probe referred to the overall number of hours spent on housekeeping activities. In the survey question examples of relevant housekeeping activities were listed. (This was not the case in the original question, which was tested in the questionnaire laboratory. In the laboratory test, the original question had generated a great variety in activities mentioned by people. So, examples were added to the question (Dehue, 1996).)

In the field test, the probe was put at the end of the base module, long after the question itself (to abandon memory effects). It elicited respondents to specify the number of hours spent per day on each of the activities mentioned as an example in the question. From these data a total number of hours spent on housekeeping was calculated. With regard to the survey question, an average number of hours was calculated. These numbers were distracted from each other, which resulted in the figures in table 8.12.

Table 8.12. Number of respondents for hours spent on housekeeping activities

Difference in number of hours spent on housekeeping activities per day		Number of respondents
hours per day	more hours measured by	
≥ 3	survey question	8 (2 %)
2		7 (2 %)
1		32 (8 %)
0	(equal)	30 (7 %)
-1		139 (33 %)
-2		98 (23 %)
≤ -3	probe	110 (26 %)
Total		424 (100 %)

These results suggest an underestimation of the overall number of hours spent on housekeeping activities as measured by the survey question. For about 80% of the 424 respondents the overall number of hours was less than the sum of the hours spent on each of the activities: for one-third the difference was one hour each day, for about half of the respondents the maximum difference was two hours each day.

An explanation may be that some activities, although included in the question as an example, are overlooked. Also it may be difficult to come up with a total number at once: the overall number is based on an unspecified rough guess. These explanations seem to be supported by comments made by respondents and registered by the interviewers. These comments indicate that the total number of hours is "estimated", "one doesn't know how many that is", and "probably too few hours have been reported". Furthermore, the comments indicate that people find it hard to come up with an answer immediately, without proper preparations.

The results on both questions show that, as expected, these are difficult questions with regard to the question-and-answer process: interpretations of the key concepts differ, it is hard to retrieve the needed information from memory and to evaluate the information. To answer the question on *'rooms'*, one has to go around the house in one's mind, see whether the rooms fit the description, and count the right ones. As for the question on *'housekeeping activities'* one has to recall one's weekly time schedule and search for housekeeping activities, the number of hours spent on these activities and add those together. However, with the alternative survey questions these problems may be less apparent than with the original ones in which no examples were included. The new formulations were felt to be an improvement with regard to validity and reliability over the original ones. By adding examples to the question (cues: Converse and Presser, 1986; cued recall: Sudman, Bradburn & Schwarz, 1996), the key concept is illustrated (making interpretations more alike among respondents), the memory search will be stimulated (by associations), and since the question is longer, the respondent is given more time to think about the answer (reducing the time pressure to answer). However, a drawback may be that cues not listed are completely forgotten because respondents are directed by the listed cues. This only holds for the question on housekeeping activities, since not all activities can be listed.

8.3.3. Discussion and conclusions

8.3.3.1. CAPI versus CATI

In the subsections above we discussed the results of the field test with regard to response, respondent and interviewer preference for one of both scenario's, and quality assessment. Here, we will summarise and discuss these results with regard to the administration mode.

As for the response, the CAPI-CATI scenario resulted in much higher rates than the CATI-CAPI scenario. A response rate of as high as 66% for CAPI-CATI is substantially more than the average response rates in former years in the three regions for the individual surveys of which POLS is composed. Thus, the CAPI-CATI scenario has been chosen. Vousten and De Heer (1998) describe the response results of the first year of POLS. The mixed-mode design resulted in 1997 in an initial response rate of 61% for base and specific module using CAPI in the first round. The use of CATI resulted in an additional 6% response for the base module, and an increase of the overall response with 2% due to CAPI re-approach. Thus, in 1997 the overall response rate was 63%,

which is a substantial increase compared to the response rates of the old surveys in 1996, being 55%. This is a remarkable result, since in the Netherlands response rates for socio-cultural surveys using CAPI have decreased in the last decade to a response level that varies between 50% and 60% (De Heer, 1999; Bakker & Winkels, 1998a).

We have seen that the debriefing questions did not reveal a clear preference for one of both scenarios neither for respondents nor for interviewers. During the debriefing sessions however, interviewers reported a clear preference for the CAPI-CATI scenario. They mentioned three reasons. The first reason (on better refusal conversion and gaining co-operation in the field than on the phone) is also discussed by Morton-Williams (1993, p. 165). She states that in face-to-face contact interviewers can prolong the interaction and may use doorstep strategies like "smiling, and maintaining eye-contact, admiring the cat, and so on." And furthermore, it is easier for people who are going to refuse anyway to do so on the phone than at the door "because the rules of politeness governing telephone interactions are less stringent." So, here again we find that the CAPI-CATI scenario has been the better choice as to one of the goals of POLS: to reduce non-response.

To see whether there are mode effects with regard to reactions of respondents, χ^2 tests have been applied to tables 8.5 - 8.11 with reaction codes. In only two of these tables we have to reject the hypotheses of independence at a 5% level: *'Non-smoking'* and *'Contacted family members'*. This may be due to the check-effect of *'Non-smoking'*, and the long *'Contact'* question and the unclear terms used. For the other questions we may conclude that respondents reacted in the same way for both modes. So, for these questions there is no mode effect.

In general this result is in accordance with other research, as is described in his review by Snijkers (1992, chapter 2). In several experiments no mode effects have been found. However, to get such a result conditions like good questionnaire design and interviewing training are essential. The need for good interviewer training has already been discussed in subsection 8.2.6. As for the questionnaire, Clark and Schober (1992, p. 36) state in their review that respondents react differently on the phone than in a face-to-face interview. "In face-to-face interviews, respondents have the full range of verbal and non-verbal signals at their disposal. (...) On the telephone they are more limited in their signals, so pauses are more disruptive. (...) Also time pressure counts. Telephone interviews go more quickly than face-to-face interviews. (...) On the telephone people give shorter answers to open-ended questions, and they pause less (...), as they were under more pressure to answer questions quickly". Thus, the questionnaire has to be adapted to the mode used.

So to conclude, CATI and CAPI are alternative administration modes, although the use of the telephone demands simple and short questions, with instant answering (no recall and lookup questions, no long lists of response options) and limited data entry (few open-ended questions). This brings us back to the questions in the questionnaire.

8.3.3.2 Quality assessment

As described above, the quality of nine questions have been investigated using reaction coding and elaboration probing. Here we will discuss these techniques and the results, and come to conclusions with regard to the quality assessment.

We have seen that reaction coding is a technique with which interviewers are able to report problems in the question-and answer process. However, this technique does not reveal the causes of the reported problems. It does not give a hint as to why respondents react in this way. Therefore, it is hard to interpret the codes. To make correct interpretations information from other sources is necessary, like in our case from the previous laboratory study, interviewer comments during the fieldwork, and interviewer debriefing sessions. Also respondent debriefings, might be helpful.

Another way to get information might be a review of the literature, and look for similar results. Also one's own experience or those of colleagues with similar questions in other test programs, can bring about extra information on how to interpret the codes. However, with these last two options one has to be careful. The question-and-answer process of individual questions is sensitive to several effects in the interview process (Converse & Presser, 1986; Clark & Schober, 1992): order effect (the context of the question), wording effect (slight changes in the wording might change the interpretation), open or closed question (questions with and without response alternatives imply different perspectives). This means that conclusions for one questionnaire may not be valid for another. In general, for deeper investigations reaction coding is only worthwhile in combination with other research.

However, reaction coding and elaboration probing are easy to implement in the CAQI questionnaire, and are no problem for interviewers to apply once they are trained. Also these techniques do not interrupt the interview too much, when they are used in a limited way. So, interviews are conducted under realistic conditions. Therefore, we feel that these techniques can be very helpful as a way to assess the quality of questions. This is also pointed out by Fowler and Cannell (1996, p. 34): "An important component missing from most methodological reports is an evaluation of questions and the quality of the information they produce. Although there are well-established standards for reporting sampling errors and response rates, there are no comparable quantitative measures of response errors. We think behaviour coding has the potential to fill the gap. (...) Users of survey data lack information about the quality of the data-collection process in general and the quality of the questions in particular. Behaviour coding with its quantitative nature and its demonstrated relationship to key measures of data quality can provide indicators to readers on both subjects." The qualitative operational field test is an example of such a study. The results of the quality assessment (as described in more detail in subsection 8.3.2) are summarised in table 8.13.

In the final questionnaire (used in the survey in 1997; CBS, 1998) the wording of two questions has been changed, as a result of the test. In table 8.14 these two questions are listed. (See table 8.1

Table 8.13. POLS questions and CAQI questions: conclusions

POLS question	Aims of the test and expected reactions	Conclusions of the field test
TimeUse	Testing of answerability	Although this is a short, closed question, it may be hard for respondents to understand the meaning of the question, make up one's mind and pick one option (not exclusive and not exhaustive): Change question wording accordingly.
WantWork	Option 'would like to but cannot' added to list of categories. Testing of answerability	By adding the extra option, the list of options now seems exhaustive. But still, the question may be hard to understand and to decide on one response option: Read the response alternatives aloud.
Health	Problems with answering categories are expected.	Answering categories may not be exclusive or exhaustive, since it may be hard to make up one's mind: Change response alternatives.
Disabled	Irritated reactions of respondents and answering before end of the question are expected.	The expectations are confirmed. The question is redundant to respondents who are not disabled: Leave the question out for those who have already answered 'no' twice to the other questions on being hindered by disabilities. Response options are not exhaustive: Add 'occasionally'.
Doctor	Long introduction: It is expected that respondents request for repetition of the question and/or for clarification.	The expectations are confirmed. This question may be hard to understand. For respondents who did not contact the doctor the information is redundant. Also, it may be hard to come up with the appropriate answer. Try to make the question shorter.
NonSmoke	Superfluous question for non-smokers. It is expected that respondents get irritated.	Respondents do not get irritated, but they interrupt question reading with an answer. This indicates that this question is superfluous to non-smokers: Leave it out.
FamCont	Long question and unclear meaning of 'contact' and 'members of the family'. It is expected that respondents request for repetition of the question and/or for clarification.	These expectations are confirmed. Respondents asked for clarification (on 'contact' and 'member of the family'). Also, some respondents answered early, indicating that the question may be too long: Change question wording and try to make it shorter.
DwelSize	Difficult task and unclear meaning of 'room'.	The interpretation of 'room' and the recall task remains difficult. Cued recall with examples listed in the question seems to be an improvement over an interviewer instruction that tells what rooms *not* to be counted, and which is only to be read at request for clarification by the respondent.
HouseWrk	Difficult task and unclear meaning of 'housework'.	The interpretation of 'housework' and the recall task remains difficult. Cued recall seems easier than free recall.

for an overview of all tested questions.) In both questions, the situation referred to has been made more specific. However, these changes will probably not abolish all problems found in the test interviews.

These results are in accordance with the results of the laboratory tests in phase 2 (Dehue, 1996), and confirmed our expectations. This indicates that results from cognitive laboratory studies, may be applicable to field survey situations. This conclusion is confirmed by Willis and Schechter (1997). In three split-ballot experiments they found that predictions based on cognitive testing were supported (p. 40): "The results provide support that there is significant carry-over between cognitive laboratory and field environment." Tucker (1997, p. 70) however makes a critical note with regard to this statement, saying that "successful generalisation from the laboratory will depend upon the researcher's ability to create realistic conditions in the laboratory or, at least, take into account the differences when drawing conclusions from laboratory experiments." Such realistic fieldwork conditions are created by the use of CAQI.

Table 8.14. Changed POLS questions as a result of the test

POLS question	Tested question wording	Final question wording used in the survey *(differences in italics)*
TimeUse		(Not changed)
WantWork		(Not changed)
Health		(Not changed)
Disabled	Are you hindered in doing daily activities caused by a long lasting illness, affection or disability in other situations, like going to work or in your spare time?	Are you *as result of* a long lasting illness, affection or disability hindered in *doing leisure activities, sporting or travelling?*
Doctor		(Not changed)
NonSmoke		(Not changed)
FamCont		(Not changed)
DwelSize	How many living, sleeping and working rooms are there in your house?	How many living, sleeping and *study or* working rooms are there in your house?
HouseWrk		(Not changed)

8.4. Conclusion

For both qualitative POLS tests (laboratory and field), CAQI turned out to be a convenient way to develop and use cognitive protocols. As for pre-tests in the cognitive laboratory, CAQI seems to be efficient when the CAI questionnaire is available, more than one interviewer conducts the interviews and a larger number of respondents are being interviewed (Snijkers, 1997). CAQI appears to be particularly rewarding for obtaining a moderate standardisation of the interviews, and a quick-and-easy transition form one pre-testing step to another. Questions and their ordering can be adapted in the course of testing, as can the various cognitive interviewing techniques to be used.

The comprehensive field tests for POLS and DVF would not have been possible without CAQI as a testing technique perfectly adapted to the computerised data collection design. Our experience with CAQI is that, even though qualitative interviewing is computerised and embedded in a standardised interview, the respondent retains enough freedom to react spontaneously to the questions. The interview is not interrupted. And the CAQI instructions help the interviewer to focus on the communication with the respondent, once they are acquainted with the protocol and know how to handle the computer. CAQI proves to be an interviewer-friendly and efficient way to administer and test a questionnaire in one and the same interview under realistic fieldwork conditions.

In general, we may conclude that CAQI is a quality assessment tool that helps to improve the quality of qualitative information on the tested questions, and thus in its turn, helps to improve the quality of survey data.

References

Akkerboom, J., and F. Dehue, 1997, The Dutch Model of Data Collection Development for Official Statistics. (International Journal of Public Opinion Research, Vol. 9, No.2, pp. 126-145.)

Akkerboom, J., F. Dehue and G. Snijkers, 1998, Qualitative tests for data collection (re-) design. (In: Bakker & Winkels (eds.), Integration of Household Surveys, pp. 20-25.)

Bakker, B., and J. Winkels, 1998a, Why integration of household surveys? - Why POLS? (In: Bakker & Winkels (eds.), Integration of Household Surveys, pp. 5-7.)

Bakker, B., and J. Winkels (eds.), 1998b, Integration of Household Surveys: Design, Advantages and Methods. (Netherlands Official Statistics, Special Issue, Vol. 13, Summer 1998. Statistics Netherlands, Voorburg/Heerlen.)

Banks, R., C. Christie, J. Currall, J. Francis, P. Harris, B. Lee, J. Martin, C. Payne, and A. Westlake (eds.), 1999, … Leading Survey and Statistical Computing into the New Millennium. Proceedings of the Third ASC International Conference. (Association for Survey Computing, Chesham, Bucks, UK.)

Blyth, B. (ed.), 1998, Market Research and Information Technology. Application and Innovation. (Esomar Monograph No. 6, Amsterdam.)

Campanelli, P., 1995, Interviewer Questionnaire on Issues in Survey Participation. (Social and Community Planning Research (SCPR), London.)

CBS, 1998, POLS Questionnaire: Base module. (In Dutch: POLS BASIS 1998. Statistics Netherlands, Department of Surveys on Living Conditions, Heerlen.)

Clark, H.H., and M.F. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), Questions about Questions, pp. 15-48.)

Converse, J.M., and S. Presser, 1986, Survey Questions. Handcrafting the Standardized Questionnaire. (Quantitative Applications in the Social Sciences Series, No. 63, Sage, Beverly Hills.)

Couper, M., 1995, Contact Description Forms. (The Joint Program in Survey Methodology, University of Maryland at College Park, Michigan.)

Couper, M.P., R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, J.M. O'Reilly (eds.), 1998, Computer Assisted Survey Information Collection. (Wiley, New York.)

De Heer, W. (ed.), 1996, Continuous Survey on Living Conditions: Qualitative Test Program. (In Dutch: POLS: Kwalitatief testtraject. Internal CBS report, No. H4266-96-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

De Heer, W., 1999, International Response Trends: Results of an International Survey. (Journal of Official Statistics, Vol. 15, No. 2, pp. 129-142.)

Dehue, F., 1996, Cognitive Laboratory Test (In Dutch: Vragenlabtest. In: De Heer (ed.), Continuous Survey on Living Conditions: Qualitative Test Program, pp. 16-26.)

De Leeuw, E.D., J.J. Hox, and G. Snijkers, 1998, The effect of computer-assisted interviewing on data quality. A review. (In: Blyth, B. (ed.), Market Research and Information Technology, pp. 173-198.)

De Leeuw, E., G. Snijkers and D. Hoezen, 1997, Kwalitatieve Veldproef POLS: II. Werving, selectie en training van meta-interviewers (Qualitative Field Test POLS: II. Recruiting, selection and training of special interviewers). Internal report in Dutch, No.: H2287-97-GWM (Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

DeMaio, Th. J., and J.M. Rothgeb, 1996, Cognitive Interviewing Techniques: In the Lab and in the Field. (In: Schwarz, N., and S. Sudman (eds.), Answering Questions, pp. 177-195.)

Everaers, P., 1996, Concept Report on Continuous Survey on Living Conditions: Qualitative Test Program. (In Dutch: Concept eindrapportage POLS: Kwalitatief testtraject. Internal CBS report, No. KPE-96-012. Statistics Netherlands, Department of Socio-Cultutal Household Surveys, Heerlen.)

Fowler, F.J., and Ch. F. Cannell, 1996, Using Behavioral Coding to Identify Cognitive Problems with Survey Questions. (In: Schwarz, N., and S. Sudman (eds.), Answering Questions, pp. 15-36.)

Kuijpers, I., and J. Akkerboom, 1997, Measuring Vacancies to be Difficult to Fulfil. (In Dutch: Het meten van Moeilijk Vervulbare Vacatures. Intersectorale vergelijkbaarheid, oorzaken en oplossingen. Internal report. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Lammers, J.P.M., I.F.E. Kuijpers and J.C. Akkerboom, 1997, Assessment of Quality Criteria in Redesign of Computer-Assisted Survey Data Collection. (In: Proceedings of the invited IASS papers presented at the ISI Session, August 18-26, Istanbul, pp. 23-32.)

Morton-Williams, 1993, Interviewer Approaches. (Dartmouth, Aldershot.)

Oksenberg, L., C. Cannell and G. Kalton, 1991, New Strategies for Pre-testing Survey Questions. (Journal of Official Statistics, Vol. 7, No. 3, pp. 349-365.)

Siegel, S., and H.J. Castellan Jr, 1988, Nonparametric Statistics for the Behavioral Sciences. (McGraw-Hill, New York.)

Schwarz, N., and S. Sudman (eds.), 1996, Answering Questions. Methodology for Determining Cognitive and communicative Processes in Survey Research. (Jossey-Bass Publishers, San Francisco.)

Snijkers, G., 1992, Computer Assisted Interviewing: Telephone or Personal? A Literature Study. (In: Westlake et al. (eds.), Survey and Statistical Computing, pp. 137-146.)

Snijkers, G., 1995, Pre-tests on ECHP-questions on Daily Activities, Pension Schemes, and Education and Training. (Internal report, No. H8170-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1996, Report on Qualitative Field Test. (In Dutch: Verslag kwalitatieve veldproef. In: De Heer (ed.), Continuous Survey on Living Conditions: Qualitative Test Program, pp. 29-35.)

Snijkers, G., 1997, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. (Bulletin de Methodologie Sociologique, No. 55, pp. 93-107.)

Snijkers, G., E. De Leeuw and D. Hoezen, 1997, Kwalitatieve Veldproef POLS: I. Opzet en uitvoering meta-interviews (Qualitative Field Test POLS: I. Design and execution of special interviews). (Internal report in Dutch, No.: H2286-97-GWM Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., E. de Leeuw, D. Hoezen, and I. Kuijpers, 1999, Computer-Assisted Qualitative Interviewing: Testing and Quality Assessment of CAPI and CATI Questionnaires in the Field. (In: Banks et al. (eds.), … Leading Survey and Statistical Computing into the New Millennium, pp. 231-258.)

Stratton, A.K., and A.M. Hardy, 1996, The Need for Questionnaire Specifications in a CAPI Environment. (Paper presented at the International Conference on Computer-Assisted Survey Information Collection, InterCASIC'96, December 11-14, San Antonio, Texas.)

Sudman, S., N.M. Bradburn and N. Schwarz, 1996, Thinking about Answers. The Application of Cognitve Processes to Survey Methodology. (Jossey-Bass Publishers, San Francisco.)

Tanur, J.M. (ed.), 1992, Questions about Questions. Inquiries into the Cognitive Bases of Surveys. (Russell Sage Foundation, New York.)

Tourangeau, R., and K.A. Rasinski, 1988, Cognitive Processes Underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol.103, No. 3, pp. 299-314.)

Tucker, C., 1997, Methodological Issues Surrounding the Application of Cognitive Psychology in Survey Research. (Bulletin de Methodologie Sociologique, No. 55, June, pp. 67-92.)

Vousten, R., and W. de Heer, 1998, Reducing non-response: the POLS fieldwork design. (In: Bakker & Winkels (eds.), Integration of Household Surveys, pp. 16-19.)

Westlake, A., R. Banks, C. Payne and T. Orchard (eds.), 1992, Survey and Statistical Computing. (North-Holland, Amsterdam.)

Willis, G., and S. Schechter, 1997, Evaluation of Cognitive Interviewing Techniques: Do the Results Generalize to the Field? (Bulletin de Methodologie Sociologique, No. 55, June, pp. 40-66.)

Wojcik, M.S., and E. Hunt, 1998, Training Field Interviewers to Use Computers: Past, Present, and Future Trends. (In: Couper et al. (eds.), Computer Assisted Survey Information Collection, pp. 331-349.)

Chapter 9

The Role of a Cognitive Laboratory within Total/Tailored Survey Design and Total Quality Management

Case Study 4: Testing The Annual Establishment Production Survey.

The Best of Two Worlds:
Total Design Method and New Kontiv Design.
An Operational Model to Improve Respondent Co-operation

Ger Snijkers and Martin Luppes

Summary: *In 1999 a program was started at Statistics Netherlands to generate ideas on how to increase response rates for self-administered business surveys by improving the communication with the respondent. The traditional approaches based on simple communication strategies (one stimulus for all units at the same moment, traditional reminder approaches using authority principles) give serious problems with response time, net response and response quality. The goal of the new program is to develop measures that change the traditional, formal and passive contact strategies into active, respondent-driven and motivational approaches. In the past several measures have already been implemented, like a cognitive laboratory to improve the wording of advance letters and questionnaires and to reduce response burden, and a guide for form-design standards to improve the layout of letters and questionnaires.*

This program is based on the philosophies of Dillman (Total Design Method) and Brög (New Kontiv Design). Where the NKD design allows for undefined respondent behaviour, the TDM approach is based on the more general standardised survey approach of well-defined respondent behaviour. In fact, both TDM and NKD are quite similar in their (respondent-driven) paradigms, but quite different in their operational approach. Both fit within Total Quality Management and have proven their benefits. Therefore the challenge is to create strategies and tactics that incorporate the best of both perspectives.

In this chapter cognitive laboratory research is placed within a broader view on Total/Tailored Survey Design and Total Quality Management. Within this context, a focus group study will be presented. With this study a redesigned form of a business survey has been pre-tested: the Annual Establishment Production Survey. While in the preceding chapters interviewer-administered questionnaires for persons and households were tested, here we have an example of applying cognitive methods to a business survey with a self-completion form.

Key Words: *Response Burden, Business Surveys, Communication Strategy*

- Paper presented at the Second International Conference on Establishment Surveys (ICES-II), June 17-21, 2000, Buffalo, New York.
- Paper published in: American Statistical Association, 2000, Proceedings of ICES-II, Invited Papers, pp. 361-371. (Alexandria, Virginia.)
- Paper published in: Netherlands Official Statistics, Winter 2000, Vol. 15, No. 4, pp. 4-10.

9.1. Introduction

Statistics Netherlands has carried out several projects to improve different parts of the communication strategies in self-administered establishment surveys[1]. These projects are aimed at improving the quality of advance letters and the effects of active strategies on response time and at the reduction of reminder calls. The success of the New Kontiv Design (Brög, 1997) in the Dutch Mobility Survey (see subsection 9.3.2), which is a household survey using an active, respondent-oriented communication strategy to raise the low response rates[2], triggered the idea of using similar approaches and tactics in establishment surveys.

Communication with respondents includes many closely interrelated aspects. However much effort is put into the planning of the survey process, the sampling and the standardisation of the questionnaires, if not enough attention is given to the position of the respondent, the form and design of the measurement instruments to be used and (most of all) what is needed to motivate respondents, the result will be low response rates and a relatively high number of time-consuming reminder calls (Dillman, 1978; Brög, 1997; Moritz & Brög, 1999). A good communication strategy should incorporate all these aspects, thus resulting in a consistent set of recommendations, procedures and instruments.

The goal of active respondent communication is twofold. First, seen from the respondent's point of view costs, time, effort and number of people involved have to be minimised. But seen from the point of view of our stakeholders and customers statistical information should be adequate, accurate, and delivered in time. It is clear that there is a tension between the customer demand and respondent's willingness. Therefore we have developed a respondent-oriented approach in which both interests are represented and balanced, based on the Total Design Method (Dillman, 1978) and the New Kontiv Design (Brög, 1997).

In this chapter we present the outline of this respondent-oriented communication strategy for business surveys. This strategy is based upon the insights acquired from several projects in the area of establishment surveys in the past years, as well as the work done in the area of social surveys. In section 9.2 we give a short review on the assumptions of the traditional and modern respondent approaches in surveys. Section 9.3 presents some results of case studies on respondent-oriented communication. Based on these results, as well as the TDM and the NKD approach, in section 9.4

[1] A communication strategy in general sense is defined as performing activities and using appropriate means, necessary within a survey to get an accurate response from a respondent within a predetermined time.

[2] In the Netherlands, non-response is a major problem in survey research, both in household and business surveys. Although business surveys are compulsory, the response rates of these surveys are considered to be too low. For example, the response rates of Annual Establishment Production Surveys (including number of employees, turnover, revenues and costs) vary between 50 and 85%. Low response rates do not only increase confidence intervals, which could lead to rather meaningless point estimates, but also cause non-response bias in the estimates (Groves, 1989; Groves and Couper, 1998). Furthermore, the response time, i.e. the time between sending out and receiving completed questionnaires is quite long, which effects the timeliness of the statistical information.

we give an outline of measures to be taken to improve survey participation in business surveys. In section 9.5 we present conclusions and some issues for further discussion.

9.2. The assumptions of survey design reviewed

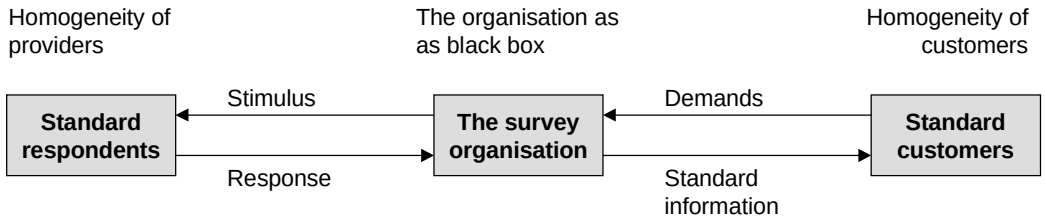
Traditionally, survey design follows the 'one-size-fits-all' strategy, which means that one design is used with one instrument for the whole sample. This approach is based on the idea that in order to get comparable response over the units, the units should receive a standardised stimulus, i.e. a stimulus in which environmental factors are controlled for. Standardisation – and not the respondent – is central in the approach, and by the same token the orientation is on the process. The respondent is in fact a 'standard' respondent (all respondents can be approached in similar standardised ways without taking into account their specific situations), who is sensitive to authority. Elements of this passive approach are: mandatory participation, based on authority and tradition, one – usually paper – questionnaire for all respondents, sent out at the same time, and no contact with respondents until the deadline has passed.

On the other side, the dissemination of statistical information was also based on the 'one-size-fits-all' approach. Customers of statistical information are being served with standard publications. Attention for demand and wishes of customers is only from very recent date (Kavaliunas & Luppes, 1998). The output of the statistical process was not adapted to that demand. Traditionally, the survey organisation sets the standards, both with regard to the input and the output. This situation is represented in figure 9.1.

The 'one-size-fits-all' survey approach has been improved by the Total Design Method introduced by Dillman in 1978. The TDM approach is based on the premises of social exchange theory in which the compliance principle of reciprocity is used, and not the principle of authority (Groves, Cialdini & Couper, 1992). The TDM assumed that maximum response could be generated by rewarding the respondent, reducing the costs for the respondent, and by establishing a relationship built on trust. This could be achieved by a coherent system using standardised, well-tested questionnaires, and an appropriate and user-friendly design (including the advance letters, brochures and other enclosures), based on carefully worded and respectfully formulated requests and instructions. This approach was applied to both mail and telephone interviews. Although this approach is respondent oriented, the basic assumption still was the 'one-size-fits-all' strategy.

In the modern situation, the survey organisation is no longer in the position to set the standards; it has to adapt to the requirements set by the environment. The modern society is characterised by a growing heterogeneity in organisations (and also in respondents). Shifting cultures create a challenge for survey organisations: organisations, both as customers of statistical information and as respondents, are no longer sensitive to authority, they have matured. Customers are no longer

Figure 9.1. The traditional survey organisation



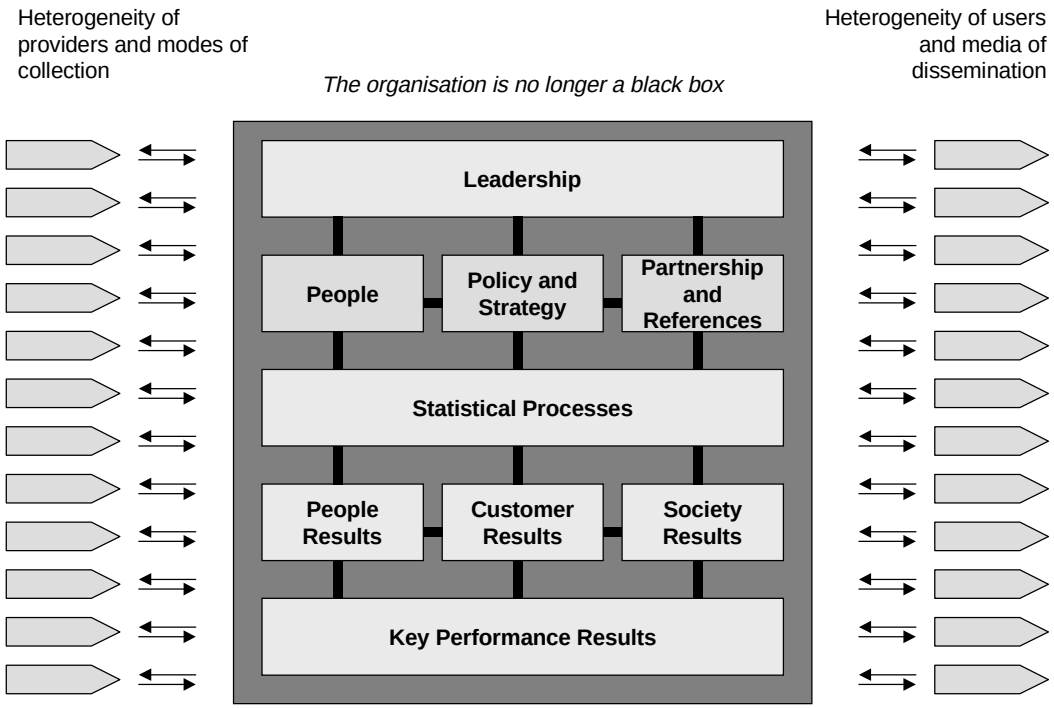
satisfied with one set of publications: they demand tailored information, both with regard to content and mode of dissemination.

Because of large numbers of surveys, there is a strong call for reduction of response burden. The consequence for the survey organisation is that the focus shifts from the survey process to the position and circumstances in which respondents respond, regardless of whether the survey population is the business or the household sector. Nowadays the general policy focuses on reducing response burden, both actual and perceived, by minimising costs, time, effort and the number of people involved. At the same time the respondents' behaviour should be influenced in such a way that an accurate response is received, i.e. all relevant data are provided correctly and on time. This means that the conditions under which the respondent has to provide information should become the focal point in the survey strategy.

The respondent-oriented survey design has been developed by Brög in his New Kontiv Design (see e.g. Moritz & Brög, 1999). The starting point for Brög is: "The researcher must adjust to the respondents, not the respondents to the researcher." In his latest adjustment of the Total Design Method, Dillman also focuses on the respondent in this way, still using 'TDM', but now referring to it as Tailored Design Method (Dillman, 2000). These approaches fit in now generally accepted ideas about Total Quality Management (TQM). The basic principle of TQM is that the quality of the product or service is not only based on internal process control, but on control of all aspects that influence the outcome of a production process[3]. The modern situation is presented in figure 9.2.

[3] Based on the ideas of Total Quality Management, the European Foundation for Quality Management introduced the EFQM model. This model is a non-prescriptive framework of nine areas for which to assess an organisation's performance: Leadership, People, Policy and Strategy, Partnership and References, Processes, People Results, Customer Results, Society Results, and Key Performance Results. The first five areas are 'Enablers', the other four are 'Results'. The 'Enablers' cover what an organisation does, the 'Results' what an organisation achieves. 'Results' are caused by 'Enablers'. The 'Enablers' learn and improve by evaluation of the 'Results'. In figure 9.2 the EFQM model (slightly adapted from the original model: turned on its side) is chosen as a model to describe the modern, open organisation.

Figure 9.2. The modern survey organisation



In order to get the best results, i.e. minimum respondent burden and maximum customer satisfaction, it is necessary to include different aspects of the survey organisation in the communication strategies. In the specific situation of communication strategies in business surveys, this means:

- Establishing a relationship with providers on the basis of mutual respect and trust.
- Motivating them to participate by explaining why their response is important and by rewarding their co-operation (using incentives). Also show that you, as a survey organisation, feel that their participation is important.
- Making partnerships with intermediate organisations like umbrella organisations, branch organisations and so on, whose interest in accurate statistical information in their field makes them receptive for motivating their members to provide information.
- Identifying subgroups of providers. Use relevant information on conditions and circumstances of the respondent in order to tailor the survey design: customise questionnaires, ask questions that can be answered easily, pick the right moment of survey, offer the appropriate mode of collection.

- Asking only relevant information, i.e. information that is actually requested by customers, and ask this information only once for the specific time base of the survey[4].
- At the same time investing in the use of registers instead of primary surveys.
- Giving relevant feedback on the performance of the organisation (many users of statistical information in establishments are also providers of statistical information).

As for the output of the statistical process, quality of statistical information (survey estimates, figures, and so on) was traditionally defined in methodological terms, like reliability, validity, measurement error, sampling bias, and so on, in accordance with the 'one-size-fits-all' approach. Less attention was given to aspects like timeliness of information, consistency of data sources, comparability with other types of information, accessibility, and so on. The call for a reduction of response burden coincides with doubts about the necessity of certain types of information and the request for new sorts of information (primarily based on integration of existing information, using the high-end functionality of ICT). This change of perspective resulted in a more balanced set of quality indicators. Overlooking the general discussions at Statistics Netherlands in past years, three quality indicators can be distinguished which determine the overall quality of statistical information (cf. Dippo, 1997; Colledge & March, 1997):

- Accuracy of the information: minimal mean square error in the statistical information. Or in non-technical terms: is the information plausible, valid and reliable?
- Timeliness of the statistical information: information about recent events has to be available as quick as possible.
- Relevance of the statistical information: information has to be of some value for customers or users.

With respect to the accuracy of the information, the existing standard statistical procedures, methods and models, such as random sampling and multivariate estimation procedures are applied to ensure minimal mean square error. Nevertheless, some assumptions underpinning the general practices are being questioned here. The basic question in surveys is whether stimuli, varying according to specific circumstances of respondents can generate the desired same response? Or in other words: is it necessary to treat every respondent with the same procedure(s) and instruments, in order to minimise response error? Or is response error a consequence of neglecting the specific circumstances in which respondents have to answer questions? During the 1970's research was initiated to increase the validity of survey data, resulting in the CASM movement (Cognitive Aspects of Survey Methodology; Jabine et al., 1984; Hippler et al., 1987). The present insight is that adapting questionnaires (wording of questions, instructions) and interview strategies (clear description of interviewer and/or respondent roles) to specific circumstances in which respondents

[4] In establishment surveys double questionnaires and/or overlap in questions do occur. The complaint, voiced by many of our providers of information, was one of the reasons statistical processes are being reorganised at Statistics Netherlands.

have to answer questions will lead to better survey results (Tanur, 1992; Morton-Williams, 1993; Sudman et al., 1996; Groves & Couper, 1998; Dillman, 2000; Sirken et al., 1999).

Plausibility of statistical information refers to its external validity. The consistency of time-series is particularly important in this respect. Macro-economic figures may vary substantially over time. Whenever there is a major change in these figures, there has to be a plausible explanation. If the researcher fails to find an explanation, this might be an indication of errors or failures in data-collection, entry or analysis. Non-response, in particular non-response that varies over time and over subgroups (resulting in selective response), complicates the assessment of the plausibility of statistical information.

Timeliness of information is very important. Practically every customer satisfaction survey on the quality of statistical information mentions timeliness as one of the most important dimensions. Statistics Netherlands uses the so-called one-to-one rule, which means that statistical information has to be published within the time period following the time period for which the survey took place. For example, figures from a monthly survey in month T have to be published in month T+1, figures from a quarterly survey in quarter T have to be published in quarter T+1, and so on. Questionnaires received after closing dates of surveys are processed at a later stage (publication of definite figures), or not processed at all. Respondents who return questionnaires too late cannot be considered as refusals or as non-respondents because they do not refuse and they do respond. Nevertheless, their response behaviour does have a negative effect on the quality of statistical information.

It is merely stating the obvious that statistical information has to be of some relevance for customers or users. In many countries, the national statistical institutes (NSI) have some form of programme, which prevents collection and dissemination of irrelevant information. For purposes of fine tuning, ongoing customer satisfaction surveys, focus groups with stakeholders and shareholders, and last but not least analysis of general trends and developments generate the necessary specifications for concepts, definitions and variables (Kavaliunas & Luppes, 1998).

High response rates have a positive effect on these quality indicators. Accuracy of information is improved by smaller variances and, given valid instruments and procedures, minimal bias. Timeliness of information is improved because more response is received within the same units of time. The basis of high response rates in establishment surveys is an efficient and active communication with providers of information[5], which takes into account the specific circumstances of the provider and his or her organisation.

[5] In the case of establishment surveys we prefer to speak of providers of information instead of respondents. The providers are the actual persons providing the information whereas the respondent is the business unit or establishment about which data are collected.

9.3. Case studies of respondent-oriented communication

9.3.1. Introduction

As we pointed out in section 9.2, respondent-oriented communication targets three dimensions of response. First of all, the response rate: as much as possible respondents should respond. This is influenced by how respondents are requested to participate. Furthermore, – and this is the second aspect: timeliness – if it is to be of any value for users, information must be made available in time. This means that the communication strategy should also optimise, or to be more precise, reduce response time (time between sending out the questionnaire and receiving it back). The third dimension is the accuracy of information, in which non-sampling error plays an important role (validity, and selectivity of response). All aspects of the means and ways of communication with the respondent are reflected in the communication strategy and there is just one moment in time where the respondent decides whether or not to participate. Given the fact that non-response varies over subgroups of the population, mixed modes of contact strategies and questionnaires are ways to minimise non-response in subgroups. However, they will increase the costs of data collection. As Groves (1989) pointed out, it helps if we know which arguments respondents take into consideration in deciding whether or not to participate. These arguments can help the survey organisation not only to redefine data-collection procedures for subsequent surveys, but also to adjust the non-response error: "…realistic models for statistical adjustment and survey administration require theories of survey participation." (ibid. p. 237).

Several projects carried out at Statistics Netherlands have addressed different theories or ideas with respect to human behaviour related to survey participation. Business surveys differ from household surveys with respect to sampling error because of substantial coverage problems, very skewed distributions within the target populations and volatile units, as well as non-sampling error, for example as a result of accessibility to relevant archives (Cox & Chinnappa, 1995). However, we believe that the non-response problem in business surveys can be described and analysed with the very same theories on survey participation as used in household surveys. In business surveys, as in household surveys (Snijkers et al., 1999), it is people that have to be persuaded to participate in the survey, and it is people that read the advance letter and have to complete the questionnaire. If these parts of the communication are not optimised (e.g. refusal conversion by addressing the right person, personalisation of the letter, and overcoming cognitive difficulties with questionnaires), a refusal to participate or to complete the questionnaire is more likely to occur (Luppes, 1998).

In this section we shall give an outline of the first approaches in this field, starting with a description of a project on the Dutch Mobility Survey, which is a household survey. The results of this project triggered the rethinking of the communication procedures in business surveys, although some initiatives had been taken in business surveys prior to this project. In this context it should be

mentioned that in 1992 the Questionnaire Laboratory was founded for cognitive testing of questionnaires (Snijkers, 1997a).

9.3.2. Project 'Improving response rate in the Mobility Survey' (1997)

At Statistics Netherlands a major impulse, with regard to respondent-oriented communication, came with the redesign of the Dutch Mobility Survey on the basis of the New Kontiv Design (NKD), developed by *Socialdata* from Munich, Germany (Brög, 1997). The Mobility Survey is held among 60,000 households. The redesign was necessary because response rates had dropped from just over 50% in 1985 to about 35% in 1998 (Moritz & Brög, 1999). The basic philosophy in the NKD is that the respondent should be regarded as a customer, to whom interviewers have to adapt all their communication instead of the other way round. In 1997 a controlled field experiment based on the NKD design was conducted in order to establish whether a significant response improvement could be achieved. This indeed proved to be the case: the response rate in the NKD sample (n=1,000) was 74%, while in the control sample the response was 44% (n=1,032).

The NKD is set up as a PAPI survey (self-completion diaries), with a telephone motivation of respondents and (possible) subsequent follow-up surveys for more detailed data in subgroups. An important advantage of this PAPI approach is that this mode of collection imposes low burden upon on the respondent. The respondents are called shortly after they have received the survey material and are motivated to complete the questionnaire and diaries. This motivation call is an important feature of the strategy. The telephone is not used to carry out the survey, but merely as an instrument to motivate the respondents.

The questionnaire itself is kept as user friendly as possible, which means as simple as possible. Basically, respondents may answer the questions in their own words, and only clearly defined and understandable categories for mode and purpose of trips are given. Pre-coded answers, explanations or definitions in the questionnaire may lead to confusion, so the design aims to put the burden of investigation on the survey organisation itself, rather than on the respondent. If the data from questionnaires are incomplete or require some clarification, additional data and information are collected by telephone. The same principles, partially structured questions and graphically well-designed (i.e. comprehensible and readable) questionnaires apply to the diaries used. The basic idea behind the diary design is to obtain information on all out-of-home activities, not only those predefined by the researcher. This leads to a quite open structure.

9.3.3. Project 'Improving advance letters of 41 Business Surveys on Annual Production' (1996)

Advance letters are quite a common feature of surveys and it is standard policy at Statistics Netherlands to send an advance letter to alert the provider or respondent of the forthcoming call,

questionnaire or interview. The quality of advance letters has often been discussed, but the lack of an appropriate theory and models lead to subjective decisions, based merely on what is considered to be appropriate. Based on the work of Cialdini (1990) and Groves, Cialdini and Couper (1992) on the compliance principles[6] underpinning requests to participate in a survey, an analytical tool was developed for content analysis of advance letters (Luppes, 1995). This tool makes it possible to describe the information content of an advance letter in relation to the psychological principles, used to convince the respondent.

This tool was applied in an internal study (unpublished) on the quality of 41 different advance letters used in Annual Establishment Production Surveys. The content analysis on these letters, performed by three independent coders, gave the following results and recommendations:

1. The enormous differences in length of the advance letters is not explained by a necessity to provide additional information that could help respondents to make a decision. On the contrary, a lot of the information, addressing definitions and detailed explanations of the survey in the longer letters probably causes confusion in stead of clarity. Extra information should be given in enclosures, not in the letter. Especially the subtle differences between anonymity and confidentiality may be diffusing (Luppes, 1994).
2. Most letters give a telephone number where respondents can obtain more information about the survey, but only 22 out of 41 letters give the name of a contact person. Personalised contacts are more effective in the provision of extra information about the survey.
3. In 31 of the 41 letters no information is given about the survey organisation (Statistics Netherlands). Eleven of the 41 letters do not address anonymity and confidentiality-related responsibility or guarantees on the part of the survey organisation. Information on the significance of the survey is given in only 9 of the 41 letters, and only 17 contain some general information about the influence survey results might have on government policy. Practically all letters address the costs of the survey in terms of information requested, the deadlines to be met and other direct costs (free return envelopes and so on). None of the letters fully informed the respondent on basic issues such as the aims and objectives of the survey, the survey organisation, the costs and benefits and the issue of anonymity and confidentiality (including informed consent issues).

9.3.4. Project 'Provider-oriented communication in Survey on Finances of Enterprises' (1998)

The traditional field strategy in the Survey on Finances of Non-financial Enterprises (SFE) was subjected to a study of whether communication with the providers could be improved. The SFE is

[6] Groves, Cialdini and Couper (1992) describe six compliance principles which can be used in the request to participate in a survey. Of these six principles the compliance principle of authority (compliance based on power difference) and the compliance principle of reciprocity (based on mutual exchange of values) are probably most common to be found in advance letters.

an annual survey among a stratified sample of enterprises with a balance sheet total of more than 25 million Dutch guilders[7]. Although response rates in the SFE are quite high (it is a compulsory survey), response time periods are quite long, many reminder calls have to be made, as well as a substantial number of calls for clarification (accuracy problems). It was hypothesised that the traditional, passive strategy of communication caused the high number of reminder calls and the long response time periods. A qualitative study was set up to investigate the following hypothesis (Oppeneer & Luppes, 1998):

Positive attention towards the provider, as expressed in (1) the advance letter based on the principle of reciprocity, (2) the moment of sending out the questionnaire depending on availability of information (annual reports of the enterprises), combined with (3) an active reminder strategy, will lead to shorter response time periods and a drop in the number of reminder calls.

A combined research strategy was used in which information of several sources was analysed:

1. quantitative description of response rates and response time per stratum;
2. content analysis of advance letters based on an adaptation of the model described in Luppes (1995);
3. focus groups with internal staff and field staff on issues of participation (especially their perception of reasons why providers would participate or not);
4. telephone interview with providers based on the results of the focus groups (their information was used to construct a topic list with respect to reasons of participation and non-participation).

Although no definite correlation between response times, number of reminder calls and the communication strategy was established, the qualitative data in the study indicate that changing the passive communication strategy into an active one will lead to a reduction of response times and the number of reminder calls. In short, the active strategy should at least consist of:

- Well-formulated advance letters, which present relevant information on the survey (purpose, costs, benefits and time to complete the questionnaire) and give the provider the sense that he or she is important for the survey. Also, thoughtful use of compliance principles is helpful in motivating the provider.
- The moment of sending out the questionnaire should be related to the availability of information. In many cases providers have no information from the annual accounts available when they receive the questionnaire. This means that the survey organisation has to keep track of the dates the information becomes available at the enterprises, and should send out the questionnaire accordingly.
- Given the fact that many enterprises are in a panel, it is helpful to use a provider's profile in which relevant dates and information are registered and which helps to customise the complete communication strategy. In fact, data-based communication approaches should be used to optimise the communication (Luppes, 1998).

[7] Information on small enterprises is collected on basis of tax registers.

9.3.5. *Project 'An efficient reminder strategy in Business Surveys on Commercial Services' (1998)*

In a qualitative study (Cörvers, 1998) on improving the reminder strategy in commercial services surveys, concepts developed in the New Kontiv Design (see project 1 in subsection 9.3.2) and the Total Design Method were used to formulate some standards and rules for improving response rates. The reason for this study were the relatively low response rates in these surveys, which could only be raised to an acceptable level by an extensive number of costly and time-consuming reminder calls, letters and duplicate questionnaires. The study resulted in the following recommendations for improving the efficiency of reminder strategy:

1. Start with questionnaires which are simple to complete, which look attractive and which use concepts and questions that respondents recognise and use;
2. Reduce the number of written reminders and use the telephone more. Although this may result in higher costs, it is more effective than reminder letters;
3. Personalise the contacts between the provider and the survey staff as much as possible. Reducing the psychological distance (i.e. establishing a relation of trust) is also possible by collecting data in a joint venture with the boards of trade. The benefits of such joint ventures make the survey more important for the providers.
4. Use an active reminder strategy in which a frequent and regular contact with providers is established. Also the quality of the contact should be high, which implies continuous training of field staff.
5. Make it clear in the case of compulsory surveys that completing and returning the questionnaire on time is a legal requirement. Just stating this requirement in the communication leads to higher response rates (Paxson, Dillman & Tarnai, 1995), although the effectiveness of this authority principle is questioned in Oppeneer and Luppes (1998).

9.3.6. *Project 'Redesign of the Annual Establishment Production Survey' (1999-2000)*

In the second half of 1998 Statistics Netherlands started a major redesign of the Annual Establishment Production Surveys. In this redesign of over 130 surveys the input, throughput and output processes are integrated into one survey, using standardised and harmonised questionnaires, consistent data collection strategies and macro-editing procedures. The redesign was triggered by a growing dissatisfaction with the relatively low response rates, the accuracy and the coherence of statistical information. Early in 1999 an extensive inventory was made of customer demands, using focus groups of stakeholders and regular customers, as well as desk research. Based on these findings, the questionnaire was redesigned early in 2000 and cognitively pre-tested in 5 focus groups with providers (Snijkers, 2000).

Although the groups were small (about 4 people per group[8]), the results were clear and in accordance with earlier results (Snijkers, 1997b):

- The advance letter did not provide an answer to all questions the providers had about the survey. For example, it did not state why the survey was conducted and what the data are used for. Providers said they would appreciate it if some results of the survey could be sent back to them (feedback).
- Providers pointed out that in the former survey, the questionnaire was sent out at the wrong time: in March, whereas the requested data become available in June. Also in June they have more time to complete the questionnaire.
- The former questionnaire asked for a lot of detailed information, which took too much time to provide. As such items were often left open, many follow-up phone calls where necessary and this probably contributed to high item non-response rates. The corresponding items in the new questionnaire were easier to answer.
- The ordering of some items in the new questionnaire was not in accordance with the administration of some businesses. For these businesses, a lot of effort was needed to obtain the requested data. Others had no problems with the ordering. In this way, subgroups of providers can be identified.
- Many providers complained that they received a large number of questionnaires every year, and, as they pointed out, they are not employed to fill in forms. What is worse, they had to provide the same kind of information for several surveys. They were not impressed by the mere fact that the surveys are mandatory.
- While at the beginning of the discussion, the providers were a bit sceptical about the aims of the meeting (improving the questionnaire), afterwards they felt it had been very useful both for them and for Statistics Netherlands. They felt they got to know Statistics Netherlands a little better.

9.4. Measures to improve respondent co-operation

Given the results of the case studies described in section 9.3, and taking into account the features of the Total Design Method and the New Kontiv Design, we define a number of measures to improve survey participation in business surveys. We present this list of measures (that may not be complete) within the design phase of the survey and data collection phase.

[8] It is difficult to organise focus groups with establishments. First, it is difficult to find the right person (i.e. the person who completed the questionnaires before), and secondly it is not easy to make an appointment. We had about 20% response for recruitment, and from those who said they would come, only about half actually showed up.

Design phase:

- *Contact person:*
  Make an effort to find the right contact person within each business, i.e. the person who has access to the requested information and is authorised to provide that information. Use all information available on past response behaviour and availability of information within the establishment to customise and personalise the communication. Databases with provider profiles form the basis for effective communication. It goes without saying that properly defining the target population and picking the sample is the first step.

- *Advance letter:*
  - Always use a personalised advance letter, well formulated and written in the right tone; make the respondent feel that he or she is important.
  - Make sure the letter is clear and short, attractive to look at and well designed, without typing errors; the tone should stimulate and motivate participation, and the language should be neutral and non-directive, without official jargon.
  - The letter should contain information on the organisation conducting the survey, the survey itself (what it is about, who has been requested to participate, an actual request to participate, including liability), why it is important to participate; the costs and benefits of participation; the issue of anonymity and confidentiality (including informed consent issues); who to contact in case of questions, i.e. name and telephone number (personalisation of contact).
  - Any additional information about the survey should be given in an extra enclosure: more information on what the survey is about and what the data are used for.

- *The questionnaire:*
  - Only ask for relevant information that cannot be collected any other way, like primary data collection using Electronic Data Interchange (EDI) from business administrations and secondary data collection using registers.
  - Only use pre-tested, well-designed and attractive questionnaires (Dillman, 1999; Jenkins & Dillman, 1997) that are user-friendly, simple to complete, without complex routing, with questions and instructions worded in such a way that they are easy to understand.
  - Use questionnaires that are tailored to subgroups. Ask each subgroup only for the information that they are able to provide. If necessary use follow-up questionnaires.

- *Mixed-mode design:*
  - Use a mixed-mode design, in which providers can choose how to provide the information, e.g. EDI, paper forms, by phone, by fax, by the Internet, or in a non-standardised way by allowing respondents to send in reports, etc. containing the requested information.
  - Only use well-tested data collection procedures and instructions.

Data collection (fieldwork):

- *Sending out the questionnaires:*
  - Make an effort to get the timing right: send out the advance letter and questionnaire at the moment the requested information is available and the respondent has time to provide that information. As many establishments participate in panel surveys, information should be available to help determine the best moment of surveying each unit. As a basic rule, for every unit the availability of annual reports should be recorded in the provider profile, together with past response behaviour.
  - Motivate providers to participate by using incentives and by making personal contact at an early stage. Phone the provider immediately after the form has been sent out in order to motivate (make an agreement about when the data can be expected, thus not giving him or her the chance to say 'no'), to show the respondent that the survey and his/her participation is important, and to maintain interaction with the respondent (building a relationship). This is essential for providers who are known to return forms late, and for those whose data are essential to the survey (e.g. large enterprises in surveys on volume of trade).
  - Try to get the data as soon as possible, after the questionnaire has been sent out. This is especially important for establishments that are essential within the survey. This makes publication of accurate estimates at an early stage possible.

- *Reminding and maintaining interaction (follow-up):*
  - Use the phone to collect missing data (reducing item non-response), immediately after the completed questionnaire has been received. Use e-mail in stead of the phone if providers prefer to communicate by e-mail. Although little is known about the real effects of e-mail communication on response behaviour, it is generally accepted that e-mail is a very efficient means of information exchange.
  - As for reminding (reducing unit non-response), switch modes: for example, use the phone, fax or the Internet in stead of mail as a first reminder. This will remind late providers more effectively to send in the requested information. Using a tailored communication strategy based on a provider profile (using all information that is known about the provider in the phone call) will increase effectiveness of the reminder strategy. In general the time between two contacts should not be too long, but the moment of the follow-up contact should always be arranged beforehand.
  - Sending back results from the survey to providers, helps to maintain interaction and to build up a relation of trust and respect. This information should preferably be of use to the individual provider (e.g. information on his line of business in his own region, or benchmark information). This is especially important in the case of panel surveys.

9.5. Conclusions

We started with a discussion on the position of providers of information and customers of that information in relation to the survey organisation in past and at present. We have seen that in the modern society the survey organisation is no longer in a position to tell providers what to do and customers what to expect as statistical products. The 'one-size-fits-all' approach is no longer appropriate, neither for the input nor for the output of the statistical process. Dillman's Total Design Method and Brög's New Kontiv Design have improved this approach on the input side of the survey process. Today, the organisation has to adapt to input and output demands, making a 'mixed-mode' approach necessary. These demands are:

- Input: reduced respondent burden and increased response rates.
- Output: increased timeliness and accuracy of relevant statistical information.

Of course, internal demands with regard to the throughput can also be put forward, such as integration and standardisation of similar processes, efficient use of statistical modelling, applying efficient sampling, stratification and weighting methods, and using other techniques to speed up the production process, e.g. macro-editing and Optical Character Recognition (OCR). As for the input and output goals, we feel that the demands can be achieved by using an active, respondent-driven and tailored communication strategy.

The respondent-oriented approach implies optimisation of the communication by using customised or tailored questionnaires and contact strategies, based on the specific conditions and circumstances of the respondent. Basically this comes down to asking the *right person* (the person who has access to the requested information and is authorised to provide that information) for the *right information* (the data that are really needed, nothing less and nothing more) at the *right moment* (when the data are available and the contact person has the time to complete the questionnaire) with the *right mode* (the mode that suits the respondent best).

However, implementation of an active, provider-oriented communication is not without consequence for the survey organisation. Apart from rising costs, and without being complete, these consequences are:

- Homogenous subgroups of providers should be identified in advance on the basis of information about which questionnaire to send and when, the preferred interviewing mode, their significance within the survey, and so on. This enables specific communication towards specific subgroups and optimises the process.
- A contact administration at the level of the provider is necessary, keeping track of all contacts with all contact persons.
- A facility for pre-testing questionnaires and data-collection modes should be present in order to optimise for questions, procedures and instructions.
- Mixed-mode designs make the statistical process much more complex since parallel processes have to be developed, thus making a detailed planning of the statistical process and

co-ordination of logistics necessary. They also have consequences for the type of statistical models used.

- A call-centre is needed, staffed by enthusiastic employees, well trained in handling different types of verbal communication strategies.

But above all, we would like to stress that putting a respondent-oriented approach into practice is more than just adapting tools and procedures. In order to be effective, the internal culture of the survey organisation should reflect the values of the modern society. Staff not only have to be trained in the use of new tools and procedures, but also in the principles of bilateral human (interviewer-respondent) interaction, in which the special relationship between a survey organisation and the respondent is reflected. Therefore, we firmly believe that the most successful survey organisations will be those which determine the perceptions, needs and wants of the sample population best, and minimise response burden through the design, delivery, and communication of appropriate and comprehensible requests.

References

Brög, W., 1997, Raising the Standard. (Keynote Paper presented at the International Conference on Transport Survey Quality and Innovation, Grainau, Germany.)

Cialdini, R.B., 1990, Deriving psychological Concepts relevant to survey Participation from the Literature on Compliance, Helping and Persuasion. (Paper presented at the 1st International workshop on Household Survey Non-response, Stockholm.)

Colledge, M., and M. March, 1997, Quality Policies, Standards, Guidelines, and Recommended Practices at National Statistical Agencies. (In: Lyberg et al. (eds.), Survey Measurement and Process Quality, pp. 501-522.)

Cörvers, F., 1998, Communication with providers in the commercial service sector: recommendations for more efficient reminding system. (In Dutch: Communicatie met berichtgevers in de commerciële dienstverlening: aanbevelingen voor efficënter rappeleren bij de sector HDV. Internal Report Statistics Netherlands, Heerlen.)

Cox, B.G., D.A. Binder, B.N. A. Christianson, M.J. Colledge, and Ph.S. Kott (eds.), 1995, Business Survey Methods. (Wiley, New York.)

Cox, B., and B.N. Chinnappa, 1995, Unique Features of Business Surveys. (In: Cox et al. (eds.), Business Survey Methods, pp. 1-17.)

Dillman, D.A., 1978, Mail and Telephone Surveys: The Total Design Method. (Wiley, New York.)

Dillman, D.A., 2000, Mail and Internet Surveys: The Tailored Design Method. (Wiley, New York.)

Dippo, C., 1997, Survey Measurement and Process Improvement: Concepts and Integration. (In: Lyberg et al. (eds.), Survey Measurement and Process Quality, pp. 457-474.)

Groves, R.M., 1989, Survey Errors and Survey Costs. (Wiley, New York.)

Groves, R.M., R.B. Cialdini and M.P. Couper, 1992, Understanding the Decision to Participate in a Survey. (Public Opinion Quarterly, Volume 56.)

Groves, R.M., and M.P. Couper, 1998, Non-response in Household Interview Surveys. (Wiley, New York.)

Hippler, H.J., N. Schwarz and S. Sudman, 1987, Social Information Processing and Survey Methodology. (Springer, New York.)

Jabine, T.B., M.L. Straf, J.M. Tanur and R. Tourangeau, 1984, Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines. (Report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology. National Academy press, Washington.)

Jenkins, C.R., and D.A. Dillman, 1997, Towards a Theory of Self-Administered Questionnaire Design. (In: Lyberg et al. (eds.), Survey Measurement and Process Quality, pp. 165-196.)

Kavaliunas, J., and M. Luppes, 1998, Measuring Customer Satisfaction: Progress and Best Practices of National Statistical Agencies. (Invited paper presented at the Joint IASS/IAOS Conference at Aguascalientes, Mexico.)

Luppes, M., 1994, Interpretation and Evaluation of Advance Letters. (Paper presented at the 5th Workshop on Household Survey Non-response, Ottawa, Canada.)

Luppes, M., 1995, A Content Analysis of Advance Letters from Expenditure Surveys of Seven Countries. (Journal of Official Statistics, Volume 11, No. 4, pp. 461-480.)

Luppes, M., 1998, Communication with Respondents and Customers, Different Goals, Similar Strategies. (Invited paper. In: Proceedings of the GSS Methodology Conference 29 June 1998, Government Statistical Service, London, UK.)

Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, D. Trewin (eds.), 1997, Survey Measurement and Process Quality. (Wiley, New York.)

Moritz, G., and W. Brög, 1999, Redesign of the Dutch Travel Survey: Response Improvement. (Paper presented at the Transportation Research Board Conference on Person Travel, Washington DC.)

Morton-Williams, J., 1993, Interviewer Approaches. (Dartmouth Publishing Company, Aldershot, UK.)

Oppeneer, M., and M. Luppes, 1998, Improving Communication with Providers of Statistical Information in Business Surveys. (In: Proceedings of the GSS Methodology Conference, 29 June 1998, Government Statistical Service, London, UK.)

Paxson, M.C., D.A. Dillman and J. Tarnai, 1995, Improving Response to Business Mail Surveys. (In: Cox et al. (eds.), Business Survey Methods, pp. 303-316).

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur and R. Tourangeau (eds.), 1999, Cognition and Survey Research. (Wiley, New York)

Snijkers, G., 1997a, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. (Bulletin de Methodologie Sociologique, No. 55, pp. 93-107.)

Snijkers, G., 1997b, Redesign of Forms and Introduction of the Euro: Focus Groups with Providers and Business Interviewers (In Dutch: Herontwerp Enquêteformulieren en invoering euro: Focusgroepen met berichtgevers en buitendienstmedewerkers. Internal CBS report. Statistics Netherlands, Division of Data Collection, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 2000, Final Report on Focus Groups on the new Form of the Annual Establishment Production Survey. (In Dutch: Eindverslag focusgroepen met nieuw concept omslagvel PS. Internal CBS report. Statistics Netherlands, Division of Data Collection, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., J. Hox and E.D. de Leeuw, 1999, Interviewers' Tactics for Fighting Survey Non-response. (Journal of Official Statistics, Vol. 15, No. 2, pp. 185-198.)

Sudman, S., N.M. Bradburn and N. Schwarz, 1996, Thinking about Answers. The Application of Cognitive processes to Survey Methodology. (Jossey-Bass, San Francisco.)

Tanur, J.T., 1992, Questions about Questions. Inquiries into the Cognitive Bases of Surveys. (Russell Sage Foundation, New York.).

Chapter 10
Conclusions

On Pre-testing, Validating Computerised Questionnaires, and Improving Data Quality

Summary: *This last chapter provides an overview of this thesis, including a summary of applied methods and results of the pre-test studies. Also aspects of criticism will be discussed, as well as strategies for presentation and gaining acceptance of pre-test research and its results. To complete this thesis, tests addressing the programming of computerised questionnaire will be described. Coming to a final conclusion, ideas for future research will end this thesis.*

Keywords: *Cognitive Research, Question-and-Answer Process, Pre-test Findings, Question Design Principles, Presentation and Acceptance of Pre-test research, Testing Methods for Computer-Assisted Interviewing, Future Research.*

10.1. Introduction

This chapter concludes the thesis. In chapter 1 I discussed the movement on Cognitive Aspects of Survey Methodology and the CASM paradigm, including the history of the Questionnaire Laboratory at Statistics Netherlands. Chapters 2 and 3 provided the setting of cognitive research at Statistics Netherlands: computer-assisted interviewing (CAI). Chapters 4 and 5 discussed the pre-test methods used at the Questionnaire Laboratory: cognitive laboratory methods and Computer-Assisted Qualitative Interviewing (CAQI). Four pre-test studies, in which these methods have been used, were discussed in the chapters 6, 7, 8 and 9. Sections 10.2 and 10.3 of this chapter summarise the methods and the results of the pre-test studies. In section 10.3, the identified problems will be related to design errors in the questions, according to question design principles.

Although pre-testing is a way to evaluate questionnaires and control for measurement errors, in the practice of survey design, its results – including recommendations for improving the questionnaire – are not always accepted. In section 10.4 arguments against pre-test research will be discussed, as well as strategies for presentation and gaining acceptance of this kind of research and its results. One of these strategies is the application of pre-test methods according to scientific principles. To complete this thesis, improvement of computer-assisted survey instruments will be discussed in section 10.5. In this section a limited number of tests will be described, addressing the programming process of computerised questionnaires and human-machine interaction. Thus, 'validating' the questionnaire with respect to these issues. I will end this thesis in section 10.6 with a discussion on the purpose of the thesis and future research, coming to a final conclusion.

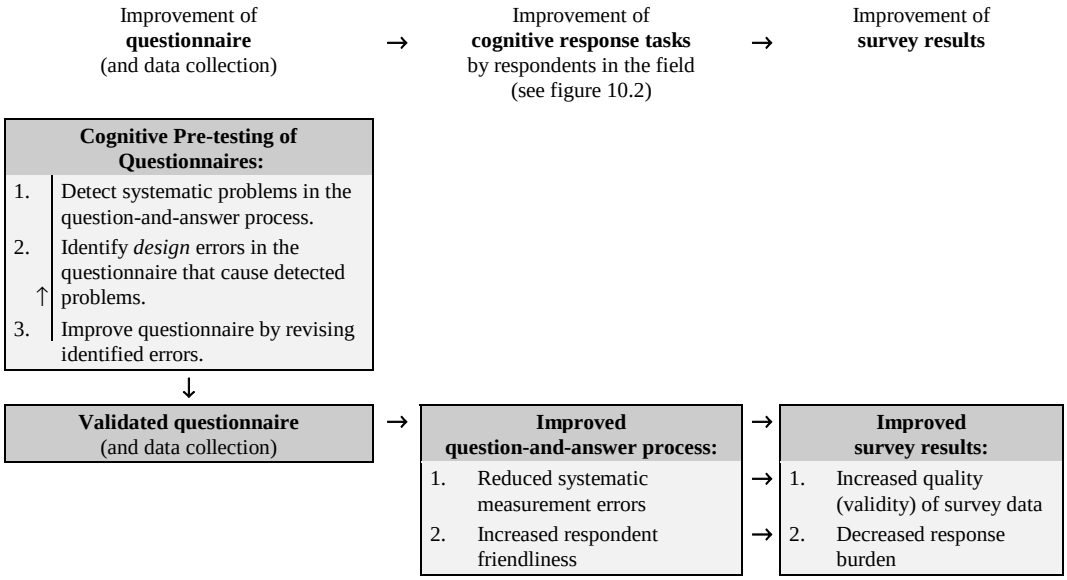
10.2. Summary of the thesis: Methods to detect problems in the question-and-answer process

(Pre-)testing questionnaires in a cognitive laboratory has been the central issue of this thesis. In chapter 1, we have seen that applied cognitive research in cognitive laboratories originated from the movement on Cognitive Aspects of Survey Methodology (CASM). Increasing the validity of survey data, by reducing measurement errors, and thus increasing the validity of derived conclusions, has been the central objective of this movement. Measurement errors are determined by a number of aspects of the survey design with regard to data collection: the questionnaire, the interviewer, the respondent, the interviewing mode, the interviewing technique, and the interview situation (see chapter 1, and chapter 2, figure 2.1). Within the context of a cognitive laboratory, the focus is on the questionnaire, the respondent and their interaction during the course of an interview.

Cognitive (pre-)testing is aimed at improving the data quality, by improving the questionnaire. By means of small-scale pre-testing the questionnaire is validated, i.e. errors in the questionnaire that cause systematic errors in the question-and-answer process of the respondent in an interview setting are detected, explained and improved (in an iterative process). In this way, the

questionnaire will be adapted to the question-and-answer process and becomes easier to answer, within a shorter period of time, and will be more respondent-friendly, resulting in reduced measurement errors, i.e. increased quality (internal validity) of survey data, and reduced respondent burden. This is the CASM paradigm (see figure 10.1), as described in chapter 1.

Figure 10.1. The CASM paradigm: Validating questionnaires and improving survey results by cognitive pre-testing

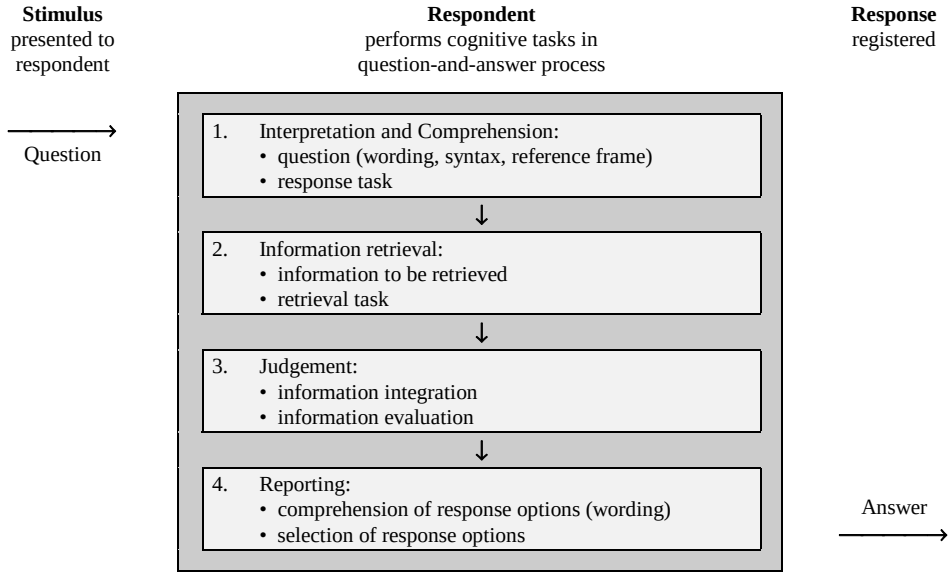


Apart from improving the questionnaire with regard to the question-and-answer process, attention has also to be given to getting the respondent to complete a questionnaire. In chapter 9 Snijkers and Luppes (2000) discussed a strategy of active respondent communication. This strategy is aimed at stimulating and motivating sampled respondents to participate in (business) surveys. In the modern society the traditional 'one-size-fits-all' strategy (one survey design with one instrument used for the whole sample) is no longer appropriate. Because of the large numbers of surveys, there is a strong call for reduction of response burden. The challenge is to influence the behaviour of the respondent in order to get an accurate response, i.e. all relevant data are provided correctly on time. The consequence for the survey organisation is that the focus shifts from the survey process to the position and circumstances in which respondents respond. The respondent-oriented approach implies optimisation of the communication by using customised or tailored questionnaires and contact strategies, based on the specific conditions and circumstances of the respondent. Basically this comes down to asking the right person for the right information at the right moment with the right mode. Within this strategy, cognitive laboratory research has been placed within a broader view on Total/Tailored Survey Design (Dillman, 1978, 2000) and Total Quality Management. The role of a cognitive laboratory is to optimise the process of administering a questionnaire, i.e. develop well-designed questionnaires that look attractive and are easy and

quickly to understand and complete. Thus, by improving the question-and-answer process, the response burden is reduced.

The question-and-answer process has been modelled by Tourangeau and Rasinski. In 1988 they presented a 4-stage model: Comprehension, Retrieval, Judgement and Reporting (see figure 10.2). In the first stage respondents have to comprehend the question, i.e. interpretation of the question wording and the task. In the ideal situation the question is interpreted in the same way as meant by the questionnaire designer. However, ambiguous wording, unclear reference sets, order effects, etc. may lead to deviant interpretations. Once the respondent has understood the question, or thinks that the question is understood correctly, he or she has to retrieve the needed information from memory or other sources to answer the question. Here, problems may arise because of difficulties in retrieving the correct information: information may be forgotten, or the question asks for specific information that is not immediately available. The third step is the formulation of an answer by integration and evaluation of information. With some questions this step is trivial, but in others a complicated calculation has to be carried out to come to an answer. The respondent may also decide not to report the true answer but to provide a socially desirable answer. After an answer is formulated, the answer has to be reported. In closed questions the judgement answer has to be mapped onto the response options. Choosing the appropriate answer may be difficult because of ambiguous wording in the items, overlapping or missing options. This model has become the basis for applied cognitive research in cognitive laboratories. (In table 10.2, a list of (violations of) design issues, in relation to this model, is presented.)

Figure 10.2. The question-and-answer process within the stimulus-response model of survey responding



Methods to research the question-and-answer process, used at the Questionnaire Laboratory, have been discussed in chapter 4. These are: expert (re-)appraisal, focus groups, in-depth interviews (including thinking aloud, follow-up probing, meaning-oriented probing, paraphrasing, targeted test questions, and vignettes), and behavioural coding. In chapter 4 these methods have been described from a practical point of view, i.e. how they are applied, thus providing an overview of current best practices.

The method that is most effective, in terms of numbers of problems detected, is in-depth interviewing in combination with thinking aloud, follow-up probing, meaning-oriented probing, and targeted test questions. This method, however, may take some time to complete (1 to 3 months). Focus groups are less time consuming (2-4 weeks), and are also very effective. The number of problems detected with expert appraisal depends on the expertise of the experts. In general, this method is less effective than in-depth interviewing and focus groups, since the questionnaire is not applied to respondents. An advantage of this method is that it can be applied very quickly (with results within 1 week). Behavioural coding is not very effective too, since it diagnoses only potential question problems, and provides no information on the source of the problems. The information gathered with this method is of quantitative nature, which makes it useful in combination with qualitative methods like in-depth interviews and focus groups. However, it takes some time and effort to get enough data. Paraphrasing and vignettes, to be used in in-depth interviews (and, as for vignettes, also in field interviews), are not effective. Therefore, these techniques are not used very often.

These methods have been presented within the context of a 5-step (pre-)test model for survey data collection development (see table 10.3 in section 10.4). In 5 steps the data collection process is developed and pre-tested, shifting from testing a first draft of the questionnaire to testing the data collection procedure in the field, and implementation of the survey. For each step, the model indicates what methods can be used. In a full pre-test program, following this model and the guidelines for application discussed in this chapter (see also section 10.3 for scientific principles of application), all aspects of the survey will be carefully tested in advance. This model offers a methodology to shift research findings from the laboratory to the field.

When starting the Questionnaire Laboratory at Statistics Netherlands in 1992, and trying to apply these methods, we were confronted with the fact that at Statistics Netherlands most questionnaires are computerised, using the interviewing program Blaise. To pre-test questionnaires for computer-assisted interviewing (CAI) in the laboratory, there were several methods to employ: (1) using a paper version of the questionnaire in which the pre-test protocol is integrated, (2) using the computer to administer the computerised questionnaire along with the pre-test protocol on paper, and (3) computerisation of pre-test protocol by integration of this protocol in the computerised survey questionnaire.

The first method was practically impossible, because of complex routing structures in CAI questionnaires. The second option was also not very practicable, since the interviewer had to integrate the survey questionnaire and the pre-test protocol during the cognitive interview. This means that the interviewer had to pay attention to two instruments, leading to omissions in the interview. This left us with option 3. We developed a method to pre-test computerised questionnaires, by integrating the pre-test protocol in the computerised questionnaire that had to be tested. We called this method Computer-Assisted Qualitative Interviewing (CAQI), as discussed in chapter 5 (Snijkers, 1997). In the questionnaire a CAQI protocol is expressed by instruction screens and probes built around the questions that are to be tested.

With CAQI, the characteristics of CAI were incorporated automatically in the pre-test instrument. These characteristics are, among others, automated routing, complex skipping patterns and branching, tailoring of questions and question wording, range and consistency checks, calculations on answers and imputations, the possibility of last minute changes, and greater standardisation of the interview. Computer-assisted interviewing, as discussed in chapter 2 (Snijkers, 1992) and in 3 (De Leeuw et al., 1998), set the conditions for pre-test research at the Questionnaire Laboratory at Statistics Netherlands. CAPI, CATI and their quality issues were the starting point for the Questionnaire Laboratory.

Chapter 2 discussed the characteristics of CAI in general, with regard to personal (CAPI) and telephone interviewing (CATI). In chapter 3 the effect of the use of the computer in interviews has been discussed. Here, we concluded that, in general, respondents are positive about the use of the computer: they attribute a greater degree of professionalism to the interview. The social interaction with the interviewer is described as comfortable and relaxed. As for the interviewers, the computer makes additional interviewer training in computer usage and computer-assisted interviewing necessary.

Now, the characteristics of CAI and the effects of the computer on the interview also hold for CAQI. Chapter 5 concluded as to CAQI: “(1) With CAQI, the cognitive protocol is conducted as it should be, without omissions. In CAQI no probes that are on individual screens are being skipped. In this way CAQI helps the interviewer to control the flow of the interview, as long as the protocol is correctly programmed. However, inadequate probing may still occur. (2) The respondent is not hindered by the computer in his reactions. (3) CAQI helps the interviewer to focus on the communication with the respondent, as long as the interviewer knows how to handle the computer and the Blaise questionnaire.” Our experience is that CAQI is a workable method and helps to conduct cognitive interviews in a standardised way.

We also concluded that CAQI results in less missing data and therefore in more information on the question-and-answer process, thus improving the quality of the information on the question-and-answer process. This conclusion was based on a small number of interviews making a comparison as to respondent and interviewer behaviour in paper-based and CAQI interviews

possible. Although the evidence on the data quality in CAQI was weak, we feel that with CAQI the data quality is at least as good as with paper-based cognitive interviews to test computerised questionnaires. In chapter 5 we also concluded that the data quality of cognitive interviews is affected by interviewer skills and instructions on the pre-test study, like adequate follow-up probing with thinking aloud.

In chapter 8, Snijkers et al. (1999) discussed the use of CAQI in a field study. In this study, our experiences with CAQI were extended. Here, we concluded that CAQI creates realistic fieldwork conditions in the laboratory, since CAQI is characterised by:

- Integration of the cognitive interviewing protocol in the standardised CAI questionnaire, thus:
  - conducting cognitive interviews in a structured way,
  - focussing on respondent's reactions.
- Reflection of standard fieldwork conditions, by using CAI and its characteristics.
- Control of the flow of the interview by the computer:
  - Resulting in appropriate probing: the appropriate cognitive interviewing techniques appropriately applied at the right time (what technique to be used when and how).
  - Thus, gathering reliable qualitative information with regard to the aim of the test program.
  - But, to achieve this:
    - every single action should be on a separate screen,
    - there should not be too much text on the screens.

And furthermore, CAQI in the field issues the possibility of:

- Greater scope of qualitative research:
  - qualitative research to be used in the field, using reaction coding, probing, vignettes and debriefing questions,
  - more data: larger number of respondents.

As for the interviewer, we concluded with regard to CAQI:

- During the interview interviewers have to pay attention to the survey questions, the cognitive protocol, the computer, and the respondent.
- To do a good job they should be well trained in:
  - properly applying (cognitive) interviewing techniques,
  - the goals (the why) of the test program,
  - handling the computer and the computerised questionnaire.

And with regard to the respondent, we stated that:

- The use of the computer in qualitative interviews gives the interviews a professional status, in which spontaneous reactions still are possible.
- The respondent had the feeling that his opinion is listened to.

The methods discussed in chapter 4 and 5 have been used in several case studies, in which questionnaires have been pre-tested. In chapter 6 (Snijkers, 1995a) a pre-test study on income questions has been discussed. Chapter 7 (Snijkers, 1995b) described a pre-test study on questions on daily activity, pension schemes, and education and training from the European Community Household Panel. In chapter 8 (Snijkers et al., 1999) a study on POLS (Continuous Survey on Living Conditions) has been described. In this study all steps of the 5-step (pre-)test model have

been followed to develop and test the questionnaire and the data collection procedures. In this chapter we focussed on step 3: the qualitative operational field test, preceding a laboratory study (in step 2). In chapter 9 (Snijkers & Luppes, 2000) an example of a focus group study has been presented, to pre-test the newly designed self-completion questionnaire of the Annual Establishment Production Survey, a business survey. The designs of these pre-test studies are summarised in table 10.1. The results of these case studies are summarised in the next section.

Table 10.1. Overview of designs of pre-test studies

Step in 5-step (pre-)test model	Chap-ter	Topics addressed in pre-test study	Design of pre-test study	
			pre-test methods used	pre-test size*
2. Qualitative laboratory test	6	Pre-testing 10 income questions	• In-depth interviews: - thinking-aloud - follow-up probes, - meaning-oriented probes - targeted test questions	• *v* = 31: - 10 self-employed - 21 employed *e* = 2
	7	Pre-testing 22 questions from the European Community Household Panel	• In-depth interviews: - thinking-aloud - follow-up probes, - meaning-oriented probes - vignettes	• *v* = 24 *e* = 4
	8	Pre-testing questions from the Continuous Survey on Living Conditions	• In-depth interviews • Focus groups	• *v* = 46 • 2 focus groups
	9	Pre-testing the newly designed self-completion form of the Annual Establishment Production Survey	• Focus groups	• *v* = 4x4 *e* = 2
3. Qualitative operational field test	8	Pre-testing the Continuous Survey on Living Conditions: • mixed-mode design • 9 questions, in addition to laboratory test (in step 2)	• Field interviews: - Behavioural (respondent reaction) coding: 7 q's - Test questions (or elaboration probe): 2 q's - Respondent debriefing questions - Interviewer debriefing questions • Interviewer debriefing sessions	• *n*(sample) = 688 *n*(response) = 365 *i*(CAPI) = 21 *i*(CATI) = 13 • *i* = (1x21), (1x13) *e* = 2

** v = volunteering respondent, n = sample respondent, e = cognitive expert/interviewer, i = field interviewer*

10.3. Pre-test study results: Question design errors

In the four pre-test studies as described in this thesis, respondents showed several difficulties with the tested questions. Table 10.2 presents an overview of these difficulties in the question-and-answer process. To summarise these findings, a list of violations of question design principles is used. This list is based on the Questionnaire Appraisal Coding System (Lessler & Forsyth, 1996; chapter 4, table 4.4), the Condensed Expert Questionnaire Appraisal Coding System (Chapter 4, table 4.5), and the eclectic classification of measurement error risks to assess questionnaires (Akkerboom & Dehue, 1997; chapter 4, table 4.6), as well as literature on question design (Oppenheim, 1992; Clark & Schober, 1992; Foddy, 1993; Brinkman, 1994; Fowler, 1995; Czaja &

Blair, 1996; ASA, 1999). For instance, Foddy's *'TAP' Paradigm for Constructing Questions* (1993, see appendix 10.1), Fowler's *Principles of Good Question Design* (1995, see appendix 10.2), and the *Key Decision Guide to Question Utility* as discussed by Czaja and Blair (1996, see appendix 10.3) are incorporated in table 10.2. The items in the list are ordered according to the 4-stage model of the question-and-answer process (Tourangeau & Rasinski, 1988; see figure 10.2), as discussed in section 10.2.

As for the income questions (in chapter 6) e.g. we concluded that the concept 'household income' was considered jargon with ambiguous interpretations. This result indicates a violation of Foddy's TAP-paradigm with regard to the 'T' (Topic) (see appendix 10.1), Fowler's principles 3a and 3b (see appendix 10.2) and Czaja and Blair's question C (see appendix 10.3), and causes problems with comprehension of the question. In order to overcome this problem we suggested to rephrase this concept as 'all incomes from everyone living here in this house', as based on respondent reactions in the interviews. Furthermore, as for comprehension problems, the question on the household income was quite long, did not end with the question itself, there was a reference set problem in connection with the prior question on the head breadwinner, and because of the reference set by monthly or yearly income other incomes were forgotten.

With respect to retrieval and judgement of information, we found that this question asked for a lot of specific and proxy information, making consultation of other sources necessary. To come to one answer a complex integration task had to be performed: adding all incomes of all persons in the household together. And as for reporting problems, respondents may report a 4-weekly income instead of the monthly income that was asked for. These findings indicate that Foddy's Applicability principle (see appendix 10.1), Fowler's principles 1d (see appendix 10.2), and Czaja and Blair's question D (see appendix 10.3) have been violated.

Now, the list in table 10.2 may not only be seen as a way to present the results, it may also be looked upon the other way round, as an operationalisation of generally accepted scientific question design principles. In this way, this list provides an independent measure for determining whether problems exist in these questions. This validates pre-test results (cf. Willis et al., 1999).

When applying this list to questions, it becomes clear where question design principles have been violated. Looking at table 10.2, it is obvious that the pre-tested questions have not been designed according to these design principles. The results from the pre-test studies, i.e. problems in the question-and-answer process, help to identify these design errors. Thus, problems in the question-and-answer process are directly related to *design* errors in the questions. Now, we may conclude that pre-test research identifies *design* errors in questions, by detecting problems in the question-and-answer process. This validates the first two steps in the CASM paradigm (figure 10.1). Once the design errors have been identified, the next step is improvement of the questionnaire. As a result of this (iterative) process, the questionnaire is adapted to the question-and-answer process: the questionnaire is validated.

Table 10.2. Overview of difficulties in pre-tested questionnaires in relation to violations of question design principles

Violation of question design principles	Questionnaire			
	Income	ECHP *	POLS **	AEPS ***
1. Comprehension of question				
• Wording:				
• technical terms, jargon	• household income • net income	• pension schemes (Qs 36-44)		X
• vague, unclear, ambiguous	• household income • net income (self-employed) • monthly/yearly income • health insurance premium • tax return in this year	• private pension scheme (Q43) • training / education (Q124)	• want paid work: paid work • contacted family doctor: telephone consultation • contacted family members: contact, member of family • dwelling size (number of rooms)	
• difficult or unclear otherwise				
• Syntax:				
• complex		• reasons for taking the course (Q127)	• time use	
• passive voice		• payment and organisation of course by employer (Q125)		
• Question:				
• long question: too many words	• net household income		• contacted family doctor • contacted family members	
• long question: long list of items	• income source	• kind of job (Q3)	• time used for house-keeping	
• several questions, double-barrelled	• health-insurance premium	• payment and organisation of course by employer (Q125)		
• double negation				
• unbalanced (one-sided); asks about causality (‘because’)		• reason for taking the course (Q127)		
• does not end with the question itself (apart from response items), but with instructions/ definitions	• net house-hold income			
• directive, non-neutral, misleading, implicit assumption		• reasons for taking the course (Q127)		
• unclear goal, redundant			• disabled • non-smoking	
• too intrusive or personal				
• hard to read aloud	• income source			
• unclear presentation, lay out				

Table 10.2 (continued). Overview of difficulties in pre-tested questionnaires in relation to violations of question design principles

Violation of question design principles	Questionnaire			
	Income	ECHP *	POLS **	AEPS ***
1. Comprehension of question (continued)				
• Reference set (frame):				
• conflict with previous question(s)	• head bread-winner → net house-hold income	• working in a job of at least 15 hours a day (Q1) → temporarily absent of job (Q2) • vocational education (Q121) → adult course (Q129)		X
• comprehension problems with key concept	• net income in ordinary month / the last 12 months			
• response items do not match question (question-answer mismatch)		• adult or language course (Q129)		
• not sufficiently specified			• contacted family doctor	
• Response task:				
• difficult				
• unclear				
2. Retrieval of information				
• Information difficult to recall/recognise:				
• question asks about hypothetical or unrealistic situation, future situation/ behaviour, complex problem	• tax return			X (complex informa-tion)
• question asks for specific information, not available by heart	• net house-hold income • tax return	• pension schemes (Qs 36-44)		X
• question refers to someone else than respondent, asks for proxy information	• net house-hold income			X
• question refers to a long period of recall				
• much information is needed to answer question	• net house-hold income			X
• Retrieval task:				
• other persons or sources or records have to be consulted to answer question	• net house-hold income • tax return			X
• questionnaire not in accordance with other sources				X

Table 10.2 (end). Overview of difficulties in pre-tested questionnaires in relation to violations of question design principles

Violation of question design principles	Questionnaire			
	Income	ECHP *	POLS **	AEPS ***
3. Judgement				
• Information integration: difficult task, difficult to come to an answer (complex calculation, estimation, guess)	• net household income		• time use • want paid work • health • time used for house-keeping	X
• Information evaluation:				
• question asks for sensitive information	• income		• non-smoking	
• risk of social desirability			• non-smoking	
4. Reporting				
• Wording of response items:				
• technical terms, jargon		• level and type of course (Q124)		
• vague or ambiguous		• level and type of course (Q124)		
• difficult or unclear otherwise			• time use	
• Response items:				
• overlapping categories		• kind of job (Q3)	• health	
• missing categories		• level and type of course (Q124) • full-time, part-time course (Q126) • language or other adult course (Q129)	• health • disabled	
• unbalanced as to distribution				
• dimensions intermingled		• level and type of course (Q124)		
• directive, non-neutral, misleading				
• long list of items		• kind of job (Q3) • level and type of course (Q124)	• time use	
• not all items are labelled				
• Response unity	• monthly, as compared to (4-) weekly income			

* ECHP: European Community Household Panel
** POLS: Continuous Survey on Living Conditions
*** AEPS: Annual Establishment Production Survey: An 'X' in this column indicates that the specific design principle is violated in the questionnaire as a whole, and with one specific question.

10.4. Improving questionnaires: Presentation and gaining acceptance of pre-test research and its results

Based on the findings as presented in the last section, recommendations for improving the questionnaire have been suggested to our clients (as we have seen in the chapters 6, 7, 8 and 9). In the cases of the income questions, these suggestions did not lead to major changes. A discussion on the pre-test results was complicated by a major reorganisation of Statistics Netherlands at the time of testing. With regard to the ECHP questionnaire, Eurostat (1995, p. 8) reported that refinements were proposed for the questionnaire, making use of the "excellent evaluation study of selected questions carried out by the Netherlands CBS for Eurostat." As we have seen in chapter 8, some POLS questions had been adapted. And, in the case of the Annual Establishment Production Survey (AEPS), the questionnaire as a whole had been adapted accordingly.

For these studies, a close collaboration existed between the cognitive laboratory and the clients, i.e. the researchers in charge of the questionnaire development. They were convinced that cognitive pre-testing is a means to detect and identify errors in the questionnaire, and that questions improved accordingly will result in better survey data. But, to make sure that changes actually are improvements, re-testing of revised questions (or when this is impossible because of e.g. time constraints, monitoring the questionnaire in the field) was recommended.

Now, we are getting into the discussion of presentation and acceptance of pre-test research and its results, including recommendations for improvement. In the literature on question design and survey methodology, pre-testing is mentioned as a way to evaluate questionnaires (investigate whether they work as intended) and control for measurement errors (i.e. assess validity) (Converse & Presser, 1986; Foddy, 1993; Fowler, 1995; Biemer & Fecso, 1995; Dippo, Chun & Sander, 1995; Czaja & Blair, 1996; Schwarz, 1997; ASA, 1999). As the American Statistical Association puts it (ASA, 1999, p. 11): "The questionnaire designer must understand the need to **pretest**, **pretest**, and then **pretest** some more." Clark and Schober (1992, p. 29) indicate why this need to pre-test: "Surveyers cannot possibly write perfect questions, self-evident to each respondent, that never need clarification. And because they cannot, the answers will often be surprising."

In the every-day practice of survey design, however, pre-testing and its results are not always accepted. Strategies for presentation and gaining acceptance of pre-test research are discussed by Rothgeb, Loomis and Hess (2000), from the US Census Bureau. The experiences and strategies they describe also hold for the Questionnaire laboratory at Statistics Netherlands.

Arguments against application of this kind of research and acceptance of its results, have been put forward by clients of the Questionnaire Laboratory at Statistics Netherlands. These arguments are also mentioned by Rothgeb et al. (2000), Fowler (1995), and Converse and presser (1986). They include:

- Basic resistance to any kind of change to the questionnaire. Sometimes researchers are opposed to any change in the questionnaire because of a deeply rooted believe that their questionnaire is well designed as it is. They do not accept any criticism what so ever.
- Concerns about trend analysis and disruption of a time series. In case of repeatedly used questions, like monthly key economic indicators, clients are particularly concerned about the impact of questionnaire revisions on time-series data. In this case, it is difficult for analysts to know whether a change in survey estimates is a true change or rather is a result of the revision of the measurement instrument. Also repeated comparison with other surveys is compromised in this way. They may argue that the questions have been used in the field without any problems. At Statistics Netherlands most surveys are continuous, meaning that, apart from major redesign programs, revision is difficult.
- Unfamiliarity with cognitive testing methods and distrust that these methods can improve the questionnaire. Most clients of the cognitive laboratory are quantitative statisticians. They are not familiar with qualitative research methods.
- Distrust of data obtained from a small non-representative sample of persons. Quantitatively oriented clients feel somewhat uncomfortable with research results that are based on a small non-representative sample of the population. They may doubt the objectivity of the interviewing and analysis process, i.e. other interviewers may obtain other data and other researchers may reach other conclusions. The data may be subject to interviewer and researcher effects, as is also pointed out by Nuyts et al. (1997).
- Unexpected recommendations for revisions. Clients may not be aware of the multitude of problems detected during testing. Also clients may be surprised by the extent of recommendations. As we have seen, recommendations may include revision of question wording or sequence, but also revision of interviewer procedures or even revision of data collection procedures.
- Improper arguments for pre-testing. In some cases testing is wanted, not with the intention of validating the questionnaire, but to get a stamp of approval or disapproval, e.g. to show to research partners that the own questionnaire is of high standards, or to show that the questionnaire of others is of low standards. When the recommendations do not fit these objectives, clients might ‘forget’ about the problems identified in the questionnaire. In these cases the cognitive laboratory is used as a ‘referee’.
- Researchers are not interested in the questionnaire but in the data. Also we find that researchers are not at all interested in the quality of the measurement instrument. In fact, they are only interested in getting data, no matter how. Only when the data show measurement errors, that undermine the survey estimates, they become interested in the questionnaire.
- Resource constraints (time, money, people) that prevent pre-testing and implementation of revisions. Time constraints are very frequently used not to pre-test questionnaires or to implement revisions. Usually there is simply not enough time to conduct even the slightest pre-test program before going into the field. Also their might be staff problems, e.g. when there is no programming staff available to implement the revisions. Sometimes money is used as an argument not to pre-test, because it would be to expensive.

A general feeling towards pre-testing is expressed by Converse and Presser (1986, pp. 51-52): "Pretesting a survey questionnaire is always recommended – no text in survey methods would speak against such hallowed scientific advice – but in practice it is probably often honored in the breach or the hurry. There is never the money nor, as deadlines loom, the time, to do enough of it. There is a corollary weakness that the practice is intuitive and informal. There are no general principles of good pretesting, no systematization of practice, no consensus about expectations, and we rarely leave records for each other. How a pretest was conducted, what investigators learned from it, how they redesigned their questionnaire on the basis of it – these matters are reported only sketchily in research reports, if at all. Not surprisingly, the power of pretests is sometimes exaggerated and their potentials often unrealized."

In addition to these remarks, Tucker (1997) discusses the scientific basis of this kind of research. According to him, in the rush to apply cognitive methods to the survey enterprise over the last decade, not enough attention has been given to scientific principles, like falsifiability, repeatability, reproducibility, and generalisability. These principles are important, in order to ensure reliability and validity of pre-test research.

According to Tucker falsifiability might be a problem since in qualitative research usually there is no theoretical direction or definable hypotheses. Repeatability and reproducibility are central to Tucker, in order to ensure the integrity of science. These principles deal with accurately describing the measurement process and secondly controlling the measurement process by standardisation. This means specifying each operation in the process, performing all operations in the same way each time, and being aware of 'interfering properties', which could contaminate the measurement at any time. As for generalizability, Tucker remarks (p. 70): "Successful generalization from the laboratory to the field will depend upon the researcher's ability to create realistic conditions in the laboratory or, at least, take into account the differences when drawing conclusions from laboratory experiments." And Tucker (p. 71) adds to this that, because of the use of small, and often convenient, samples in cognitive research, "the results from one realisation of an experiment to another might not be stable even when done by the same researcher under the same conditions."

Some of the arguments listed above are related to the lack of attention given to these principles. Thus, the first – and obvious – way to overcome problems of distrust by clients is a well-designed pre-test program, carried out according to scientific principles. In chapters 4, 5 and 8 we already discussed several principles to guarantee for reliability and validity of cognitive research. They include:

- Standardisation of methods. Methods are carried out in a standardised way as described in chapter 4. In chapter 5 (Snijkers, 1997) we have also seen that Computer-Assisted Qualitative Interviewing (CAQI) helps to standardise the execution of cognitive methods. With standardisation, procedures are repeatable and results are reproducible. In this way, interviewer and observer variance is reduced and the reliability of the data and the conclusions are improved.

- Team work. A team of researchers prepares, carries out, and analyses a cognitive study, according to the standards. And finally, together they try to reach inter-subjective or more objective conclusions. In this way, interviewer and observer bias is reduced.
- Triangulation. Several methods are applied within one study, both to get confirmation of results obtained with other methods and to get complementary results. Thus, the validity of research conclusions is controlled for.
- Use of the 5-step (pre-)test model of data collection development.
  - With regard to generalisation of results, we have seen that CAQI creates realistic fieldwork conditions for pre-testing questionnaires both in the laboratory and in the field. To get even stronger evidence on the results, the number of interviews may be doubled or even tripled. In chapter 8 (Snijkers et al., 1999), we concluded that results of cognitive laboratory studies generalise to the field. This hypothesis is also supported by a three split-ballot experiment carried out by Willis and Schechter (1997). However, to get convincing evidence on generalisation for clients, a field test (step 3 in the 5-step (pre-)test model of data collection development) is needed.
  - As for falsifiability of results, the 5-step (pre-)test model of data collection development is also very helpful. Hypotheses on the questionnaire may be formulated in one step and tested within the next.

  With the 5-step model results are tested and step-by-step generalised. This model offers a methodology to shift research findings from the laboratory to the field.

In addition to a well-designed pre-test program, the second way to overcome distrust is presentation of results in a detailed research report. Such a report should not only include a description of the methods used, but above all a detailed description of the problems identified, the recommended revisions, and justification of the revisions. Furthermore the report should include quotes from cognitive interviews to illustrate the identified problems. This is what clients are mostly interested in. They may want to check the conclusions for themselves. Examples of these reports have been presented in the chapters 6 (Income) and 7 (ECHP).

However, still clients may not accept the results of a study, for reasons listed above. To overcome problems of distrust at the end of a study, it is recommended to communicate on the study with clients at an early stage and to involve them during all stages of the study (cf. Rothgeb et al., 2000). As for the planning phase of the study, Rothgeb et al (2000) suggest to do the following[1]: (1) have an open discussion with clients about the research objectives, scope, and timeliness of the research, (2) identify the stakeholders and decision makers, (3) identify 'unchangeables' with the survey, (4) obtain documentation of question objectives from clients, (5) become familiar with subject matter issues, (6) develop a research proposal, and (7) meet with

[1] These aspects are in accordance with general aspects that have to be dealt with while planning a project: (1) identify the stakeholders and decision makers, (2) identify the origins of the project, (3) identify the targets, (4) define the result, and (5) identify the constraints.

the client and discuss this proposal, including outstanding issues, prior to the beginning of the research data collection. Although these activities may be time-consuming, we feel that they are important in getting commitment from the client, as in (1) and (7). As for the researchers, it is important to identify the constraints of the study, in (1), (2) and (3), and to get familiar with the subject, in (4) and (5), in order to develop a specific, acceptable, and realistic research proposal (6)[2], in which the scientific principles discussed above are incorporated.

During the research phase Rothgeb et al. (2000) again recommend to involve the client: (1) invite the client to review the cognitive interview protocol, (2) invite the client to observe live cognitive interviews or video recordings of interviews, and if possible (3) use a quasi-split panel design when clients insist on specific question wording. Especially inviting the client to observe live cognitive interviews (2) is very convincing. Our experience is that researchers, who were very sceptical about cognitive pre-testing of questionnaires, accepted the research results, only after seeing one respondent struggling with a questionnaire. Having clients observe these interviews and witness questionnaire problems firsthand leaves a lasting impression on them. As for (1), it is recommended to prevent clients from becoming too involved. And (3) permits the testing of both the client's version of a question and the cognitive researcher's wording. However, this is only possible with a large enough test size, and when the research constraints permit this.

After the research has been carried out, results and recommendations have to be presented to the client. A common way to do so is the preparation of a research report that includes all aspects as mentioned above. A videotape with convincing parts of the interviews may be added to the report. Apart from presenting the results and recommendations in these ways, again Rothgeb et al. (2000) recommend to meet with the client to discuss those. During the meeting, they suggest to utilise the following strategies: (1) directly connect each specific recommendation to the client's objectives and emphasise potential for improved data quality, (2) if the client demonstrates resistance, be open to compromise and try to find a solution that fits the client's objectives, (3) recognise that the client is the final decision maker and that the cognitive laboratory researchers serve in an advisory role, and (4) reissue the research report including final decisions reached for each recommendation. Although, this may seem obvious, it is often forgotten, in the rush to move on with new projects. Reports are often send to the client, without discussing them. Like at the start and during a project, meeting the client to present and discuss the results, is equally important at the end of a project in order to gain acceptance of results.

These strategies may overcome resistance because of psychological, statistical or scientific distrust. Resistance because of resource constraints (time, money, people) however, is harder to

[2] In general, a research (or project) proposal has to meet the following criteria, as abbreviated in the acronym SMART: Specific (Is the proposal specific enough?), Measurable (are the results, as described in the proposal, measurable?), Acceptable (is the proposal acceptable to everyone involved), Realistic (is the proposal realistic to everyone?), and Timeliness (can the research be carried in time?).

deal with. The most frequently used reason not to pre-test questionnaires is because of tight time schedules. Researchers are always working under heavy time pressure to get the survey ready in time. Very often researchers come to ask for advice when there is almost no time left before going into the field. In those situations there is only limited time to do an expert appraisal. Exceptions to these situations are newly designed questionnaires (like the income questions and the ECHP) and major redesign programs (like the POLS questionnaire and the Annual Establishment Production Survey). Here, it is recommended to make a research proposal, according to all principles discussed above, including a realistic planning of all stages of the development process of the survey.

With respect to costs, Fowler (1995, p. 136) puts forward that this kind of research is "extraordinary inexpensive in the context of most survey budgets. Assuming that a survey budget included some kind of pre-test, this research will have a very small percentage impact on total costs." When money constraints are tight, we suggest to reduce the survey sample, meaning that less interviews have to be conducted, in order to finance a pre-test study. As for constraints on people, the management responsible for surveys has to prioritise issues of importance. Furthermore, like at Statistics Netherlands, it is recommended to have a group of people who are dedicated to cognitive research.

In addition to the strategies discussed above, Rothgeb et al. (2000) suggest to consider establishing a Pre-testing Policy (USCB, 1998). At the U.S. Bureau of the Census such a policy has been developed, together with a pamphlet stating the policy and descriptions of the various pre-testing methods that can be used. This pamphlet was distributed to internal and external clients. According to Rothgeb et al. (p. 12), "the existence of the policy does invoke some influence when discussing pre-testing and its research results. Clients realise they cannot add or change questions without some kind of testing. (…) The existence of the policy is clearly beneficial to the Census Bureau's effort to improve survey measurement." At Statistics Netherlands, very recently quality indicators in order to monitor the survey process, including questionnaire design, have been developed (Van Berkel et al., 2001). The indicators on questionnaire design deal with violations of design aspects (as listed in table 10.1, to be used for expert appraisal), and testing the questionnaire in the cognitive laboratory and in the field. All together 20 indicators have been formulated to be used by questionnaire designers to indicate the quality of a questionnaire. These indicators are not as strong as a Pre-testing Policy, but they serve the same goal: improvement of survey measurement.

10.5. Improving computerised questionnaires: CAI (pre-)testing

Up till now we have discussed improving questionnaires by cognitive pre-testing, in order to improve the quality of survey data. However, when applying Computer-Assisted Interviewing (CAI), as discussed in chapter 2 (Snijkers, 1992) and 3 (De Leeuw et al., 1998), not only the

question-and-answer process has to be tested, but also the computerisation of the questionnaire. In chapter 2 (figure 2.1, Snijkers, 1992) we have seen that CAI design errors may also cause errors in the survey data, and that testing with regard to CAI is necessary to ensure a correctly working questionnaire program. In this section we will briefly discuss CAI testing of the questionnaire.

We have seen that the design process of a questionnaire in step 1 of our 5-step (pre-)test model starts with the research objectives and concepts to be measured, followed by operationalisations. The final result of step 1 is a prototype of a questionnaire, or, in case of CAI, specifications for a computerised questionnaire. Issues related to specification development are discussed by Kinsey and Jewell (1998), in which the attributes of CAI (as discussed in chapter 2; Snijkers, 1992) have to be considered.

On the basis of these specifications a computerised questionnaire is developed. Design issues as to CAI programming are discussed by Snijkers (1992; chapter 2), Kinsey and Jewell (1998), and Pierzchala and Manners (1998). These issues include, among others, standardised programming, modular design (with common data structures and standardised procedures with regard to question and screen formats, editing, etc.), portability (re-usability of questions from other questionnaires stored in question libraries), and version control.

The computerised questionnaire has to operate correctly in every single interview, i.e. there should be no programming bugs. Furthermore, the questionnaire should be easy to handle by interviewers and programmed according to standard screen layout conventions. And, as for the interviewing process, the computerised questionnaire has to operate correctly on the laptop computer of the interviewer, and within the administration system of data processing, i.e. sending empty records to the interviewers and getting completed records back. In addition to the computerised questionnaire a paper version is developed; the paper version (including question wording and routing structures) is not used for interviewing but for communication and documentation. Both versions have to be identical. A number of tests related to these issues are discussed by Kinsey and Jewell (1998). Here, we will focus on the most important tests: the functionality test and program inspections, the usability test, and the hardware test. It may be clear that there is some overlap between these tests.

The objectives of the functionality test are twofold. The first goal is to find out whether the computerised questionnaire works correctly in every interview, i.e. whether the questionnaire is programmed according to the specifications. This includes testing all question and answer option wordings, routings, wording variations and fills, consistency checks, and other specified instrument features. To compare the source code with the specifications, programmers may also use program inspections. Secondly, a functionality test is used to find out whether the computerised questionnaire and the paper documentation are identical.

The second test is the usability test. This test focuses on the human-machine interaction (HCI) and the user friendliness of the CAI instrument by end-users: Is the CAI instrument easy to use by interviewers (in case of CAPI or CATI) or respondents (in case of CASI)? 'Good' usability is the result of testing the entire package, including software, hardware, manuals and training, and has to be considered at every development stage. Here, issues like operating the CAI program and the computer, appropriate screen layout in relation to the data collection mode, readability of the questions, duration of an interview, clarity of instructions (in the manual, the computer or during training) are tested. According to Kinsey and Jewell (1998) usability testing requires a structured approach in a usability laboratory, similar to those offered by cognitive laboratory methods. Couper (1999, 2000) discusses several methods of usability testing, including usability inspection or HCI expert evaluation methods, end-user evaluation methods, laboratory-based observation methods, laboratory-based experiments, and field-based usability evaluation.

The last test to be described here is the hardware test. In this test the CAI instrument is tested in the production environment, with the equipment and systems used by end-users. According to Kinsey and Jewell (1998), this test ensures that system constraints (e.g. size, memory) are not violated, and the program loads and runs in a reasonable time. Additionally, the testing process should confirm that the capacity and speed of batteries, modems, hard disks, and other hardware meet the needs of the survey.

These tests may be used for pre-testing a computerised questionnaire in a laboratory setting (in step 2 of the 5-step (pre-)test model of data collection development), before the questionnaire is to be used in the field (e.g. in a qualitative field test in step 3). CAI pre-testing allows developers to simulate the CAI interviewing process and test the CAI program. Thus, according to Kinsey and Jewell (1998, pp. 119-120), CAI pre-testing includes "the ability to (1) conduct a structured functionality test, (2) evaluate usability of the instrument and any support systems, (3) collect and examine test data", and (4) identify other items that need improvement. These tests may also be used in steps 3 and 4 for field-based evaluations. Now, we can adjust the 5-step (pre-)test model of data collection (chapter 4, table 4.1) to Computer-Assisted Survey Information Collection (CASIC). In table 10.3 the CAI test methods are included in the model.

At Statistics Netherlands staff of the fieldwork department carries out these tests. They run the computerised questionnaire on their computer, following the paper documentation. They may also use pre-defined test cases (i.e. simulated respondents) to find out whether all instrument features are correct, or employ test interviews with volunteering interviewers and respondents. However, in general, these tests are not carried out as carefully as should be. Reasons for this are put forward by Kinsey and Jewell (1998). First of all, they indicate that a lack of adequate specifications makes it difficult to assess the accuracy of the program. Secondly, tight schedules and lack of staff make it difficult to test the entire program sufficiently. Furthermore, a lack of sufficient testing tools slows the testing process and reduces the efficiency of the testers. And finally, a lack of a structured testing environment, testing procedures and testing plan affects the effectiveness of the testers.

It may be clear that, as Snijkers (1992, p. 138; chapter 2) remarks, "testing of interview programs is very important to ensure that the program will operate correctly in every situation." As with cognitive pre-testing, here we may also state that CAI pre-testing 'validates' the computerised questionnaire, not with respect to the question-and-answer process, but with respect to the programming process and the human-machine interaction.

Table 10.3. A 5-step (pre-)test model of CASIC development: Cognitive and CAI test methods (*)

Step	Result of step	Topics to be addressed	(Pre-)test methods	
			Cognitive (and other)	CAI
1. Definition/ feasibility study	Prototypes of: • questionnaire, • data collection procedure • data processing procedure	Designing: • questionnaire, • data collection and • data processing procedure: What data have to be gathered how?	• Review of literature, research papers, meta-analysis • Expert appraisal • Exploratory focus groups	• Program inspection
2. Qualitative laboratory test	'Less error-prone' questionnaire, by revision of: wording, sequence of questions, inter-viewer procedures, interviewing mode	Pre-testing questionnaire: • cognitive pre-testing (question-and-answer process) • CAI pre-testing (functionality, usability, production environment requirements)	• Expert (re)appraisal • Focus groups • In-depth interviews • Observations (monitoring standardised interviews)	• Program inspection • Functionality test • Usability test (laboratory-based) • Hardware test
3. Qualitative operational field test	'Less error-prone': • questionnaire, • data collection and • data processing procedure	Pre-testing: • questionnaire (measurement quality), and • data collection and • data processing (process efficiency) in the field	• Monitoring standardised interviews: evaluation/test questions, observations • Focus groups / debriefings • Re-interviews • Expert (re)appraisal • (Item) non-response analysis	• Program inspection • Functionality test • Usability test (field-based) • Hardware test
4. Quantitative pilot study	Final questionnaire, data collection and data processing procedure	Testing data collection and data processing procedure in the field: costs and benefits	• Monitoring standardised interviews: evaluation/test questions, observations • (Item) non-response analysis • Data analysis (external validity) • Controlled statistical experiments • Other monitoring methods: re-interviews, focus groups/debriefings	• Functionality test • Usability test (field-based) • Hardware test
5. Implemen-tation	Get survey going: all preparations for carrying out the survey have been made	Implementation of final questionnaire, data collection and data processing procedure	• Monitoring standardised interviews in survey: evaluation/test questions, observations • Other monitoring methods: re-interviews, focus groups/debriefings • (Item) non-response analysis, data analysis	

(*) CASIC: Computer-Assisted Survey Information Collection.
Table adapted from: Chapter 4, table 4.1; originally adapted from: Akkerboom & Dehue (1997: table 1a, p. 129).

10.6. Conclusions and future research

In this thesis I have discussed several issues of cognitive pre-test research. I started with the history of cognitive aspects of survey methodology. Then I discussed the CASM paradigm, pre-test methods (applied within computer-assisted interviewing), examples of pre-test research, the results of these case studies, and the presentation of these results.

In this thesis I have tried to systematically describe my experiences (including those of colleagues at the Questionnaire Laboratory at Statistics Netherlands) with pre-test research. The purpose of the thesis is documentation of this practice. This text is not aimed at a theoretical discussion of cognitive methods, but at discussing the application of these methods: setting-up and carrying-out pre-test research, analysing the data and presentation of the results. As far as I know, in the literature on cognitive aspects of survey methodology, little is said about how to apply this kind of research in practice. This is confirmed by Willis et al. (1999, p. 137), who discuss systematic schemes for the description of how cognitive interviewing methods are practised. They conclude that "(…) no such scheme exists for use in cognitive interviewing research." With this text I have tried to provide such schemes by systematically describing the state of the art at the Questionnaire Laboratory at Statistics Netherlands, in order for other researchers to continue from here. I do hope I have succeeded in this goal.

Given these methods and their results, there are still a number of issues for future applied research that have my interest. These issues (some of which are also proposed by Martin and Tucker (1999)) build on the state of the art and current best practices. They focus on making pre-test research results transparent to questionnaire designers and development of lacking methods. They include:

- *Application of meta-analysis to cognitive laboratory research papers from research institutes all over the world.* This research is aimed at systematically describing the state of the art of cognitive research, with regard to methods (i.e. describing current best practices, and investigating what kind of methods identify what kind of problems), results (what kind of questions result in what kind of problems in the question-and-answer process), and recommendations for questionnaire design (what kind of recommendations are proposed to solve what kind of problems). This research should answer the question put forward by Willis et al. (1999, p. 148): "What have we learned in general about questionnaire design, based on the thousands of cognitive interviews that have been conducted, that can be used to inform the crafting of survey questions?"
- *The combination of cognitive research results with results from non-CASM approaches that address the quality of questionnaires*, like split-ballot experiments, MTMM (multi-trait multi-method) experiments (Saris, 1998) and interaction analysis (Van der Zouwen & Dijkstra, 1998). In the split-ballot experiment effects of differences in methods are investigated, like different question wordings being asked to subsamples from the same population. The researcher wants to know whether it makes a difference if one or the other question is used. In

the MTMM approach the researcher wants to know whether the method makes a difference and what the reliability of the questions is. In this approach information on the reliability and validity of questions is obtained in three traits (observations) during one survey. In interaction analysis the course of an interview is investigated, i.e. the interactions between interviewer and respondent, in order to get information on the quality of the questionnaire. These research approaches have not been discussed in this thesis. However, a combination of results from several research approaches will result in even more information on the quality of questionnaires and will help to improve the crafting of survey questions.

- *Designing a question database, that includes question wordings and meta-information on these questions.* Instead of designing new questions from scratch all over again, researchers may design questionnaires by selecting questions from this database. This database includes all information collected and coded in the above-mentioned research and is made available through the Internet. Thus, more information on the measurement instrument becomes available, making data from several studies (with the same questions on background variables) comparable, and having more indications on the quality of the data (with regard to reliability and validity).
- *Development of methods to continuously monitor the quality of questionnaires and derived data, i.e. measures of response errors*. At Statistics Netherlands most personal and household surveys are continuous. As for those surveys, and for re-use of questions in the question database, pre-testing once is not enough. As Converse and Presser (1986, p. 51) indicate: "… the meaning of questions can be affected by the context of neighboring questions in the interview." And furthermore "language constantly changes", making question wordings subject to changing interpretations. In chapter 8, I quoted Fowler & Cannell (1996), who argued that behaviour coding might be used in this way. However, more research addressing this issue is needed.
- *Development of an empirically based strategy to stimulate and motivate respondents to participate in surveys.* This research is aimed at identifying stimuli that stimulate and motivate respondents to participate in surveys, in order to optimise the contact strategy. In chapter 9, Snijkers and Luppes discussed ideas for such a strategy of active respondent communication. However, there is little evidence on what parts of the strategy are most effective. Also, as for implementation of such a strategy, information on cost and process efficiency is needed.

In this thesis the CASM paradigm has been taken for granted. In the literature on CASM, however, the effectiveness of pre-test research has often been discussed. After a questionnaire has been pre-tested and the problems have been 'fixed', several questions might be asked:

- *Did the cognitive methods really reveal systematic problems in the question-answer process?*
- *Has the questionnaire really been improved?* Are pre-tested questionnaires really better than questionnaires that have not been tested?
- *Will respondent burden decrease?* Will it be it easier for respondents to come up with an answer? Will the questionnaire be more respondent-friendly?
- *Will the survey data really be better?* Will measurement error reduce and validity increase?

Related questions have been posed by Groves (1996, pp. 401-402) when he asked himself "How do we know what we think they think is really what they think?", while discussing the usefulness of cognitive research:

"1. Is there evidence that a discovered 'problem' will exist for all members of the target population? Is evidence sought that different problems exist for groups for whom the questions are more salient, more or less threatening, more or less burdensome?
2. Do multiple measures of the same component of the question-answer technique discover the same problem (that is, exhibit convergent validity)?
3. When the problem is 'fixed', does replication of the techniques show that the problem has disappeared?
4. When the problem is fixed, does application of other techniques discover any new problem?
5 Is there evidence that the fix produces a question with less measurement error than the original one?

These kinds of questions require experimental designs with contrasts of new and old questions and explicit measures of accuracy. Such are common to survey methodology for studies of measurement error and are needed to demonstrate the validity of the techniques covered in this volume." At CASM II, Schwarz (1999, p. 71) also quoted Groves and was surprised to see that "in the light of the extensive applied work done in cognitive laboratories, (…) a systematic evaluation of the practical usefulness of cognitive laboratory procedures is still missing."

Although, this thesis is not aimed at discussing these issues, I feel that this thesis contributes to providing evidence on the effectiveness of the CASM paradigm. As for Groves' first question, in chapter 8 (on POLS, Snijkers et al., 1999) we concluded that laboratory results generalise to the field. The case studies in chapter 6, 7, 8 and 9 showed that for respondents with the same background characteristics, the same kind of problems have been discovered. The study in chapter 8 also provides an answer to questions 3 and 4: respondent reactions to questions revised after a prior laboratory study showed that the original problems were fixed, although new problems emerged. However, to get strong evidence on the issues raised in questions 3 and 4, re-testing of the revised questionnaire, including experimental studies, is needed. As for Groves' last question, and my questions above, in section 10.3 I showed that these questions might be answered positively.

Although there is still little empirical evidence that pre-test methods help to improve the questionnaire and thus improve the quality of derived survey data, the data presented in this thesis do show that pre-test research – when carried out according to scientific principles – identifies problems in the question-and-answer process for questions that do not meet scientific design standards. These data support the hypothesis that pre-testing in a cognitive laboratory validates questionnaires.

References

Akkerboom, H., and F. Dehue, 1997, The Dutch Model of Data Collection Development for Official Surveys. (International Journal of Public Opinion Research, Vol. 9, No. 2, pp. 126-145.)

ASA, 1999, Designing a Questionnaire. ASA series: What is a Survey? (American Statistical Association, Alexandria, VA.)

Banks, R., C. Christie, J. Currall, J. Francis, P. Harris, B. Lee, J. Martin, C. Payne, and A. Westlake (eds.), 1999, … Leading Survey and Statistical Computing into the New Millennium. Proceedings of the Third ASC International Conference. (Association for Survey Computing, Chesham, Bucks, UK.)

Biemer, P.P., and R.S. Fecso, 1995, Evaluating and Controlling Measurement Error in Business Surveys. (In: Cox et al., Business Survey Methods, pp. 257-281.)

Blyth, B. (ed.), 1998, Market Research and Information Technology. Application and Innovation. (Esomar Monograph, No. 6, Amsterdam.)

Brinkman, J., 1994, The Questionnaire. (In Dutch: De vragenlijst. Wolters-Noordhof, Groningen.)

Clark, H., and M. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), Questions about Questions, pp. 15-48.)

Converse, J.M., and S. Presser, 1986, Survey Questions. Handcrafting the Standardized Questionnaire. (Quantitative Applications in the Social Sciences Series, No. 63, Sage, Beverly Hills.)

Cox, B.G., D.A. Binder, B.N. A. Christianson, M.J. Colledge, and Ph.S. Kott (eds.), 1995, Business Survey Methods. (Wiley, New York.)

Couper, M.P., 1999, The Application of Cognitive Science to Computer-Assisted Interviewing. (In: Sirken et al. (eds.), 1999, Cognition and Survey Research, pp. 277-300.)

Couper, M.P., 2000, Usability Evaluation of Computer-Assisted Survey Instruments. (In: Social Science Computer Review, Vol. 18, No. 4, pp. 384-396.)

Couper, M.P., R.P. Baker, J. Bethlehem, C. Clark, J. Martin, W.L. Nicholls II, J.M. O'Reilly, 1998, Computer Assisted Survey Information Collection. (Wiley, New York.)

Czaja, R., and J. Blair, 1996, Designing Surveys. A Guide to Decisions and Evaluation. (Sage, London.)

De Leeuw, E.D., J.J. Hox, and G. Snijkers, 1998, The effect of computer-assisted interviewing on data quality. A review. (In: Blyth, B. (ed.), Market Research and Information Technology, pp. 173-198.)

Dillman, D.A., 1978, Mail and Telephone Surveys: The Total Design Method. (Wiley, New York.)

Dillman, D.A., 2000, Mail and Internet Surveys: The Tailored Design Method. (Wiley, New York.)

Dippo, C.S., Y.I. Chun, and J. Sander, 1995, Designing the Data Collection Process. (In: Cox et al., Business Survey Methods, pp. 283-301.)

Eurostat, 1995, Working Group “European Community Household Panel”: Questionnaire for Wave 3: proposed changes. (Eurostat report, doc.PAN 54/95. Statistical Office of the European Community, Luxemburg.)

Foddy, 1993, Constructing Questions for Interviews and Questionnaires. Theory and Practice in Social Research. (Cambridge University Press, Cambridge.)

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

Fowler, F.J., and Ch.F. Cannell, 1996, Using Behavioral Coding to Identify Cognitive Problems with Survey Questions. (In: Schwarz & Sudman (eds.), Answering Questions, pp. 15-36.)

Groves, R.M., 1989, Survey Errors and Survey Costs. (Wiley, New York.)

Groves, R.M., 1996, How do we know what we think they think is really what they think? (In: Schwarz & Sudman (eds.), Answering Question, pp. 389-402.)

Kinsey, S.H., and D.M. Jewell, 1998, A Systematic Approach to Instrument Development in CAI. (In: Couper et al. (eds.), Computer Assisted Survey Information Collection, pp. 105-123.)

Lessler, J.T., and B.H. Forsyth, 1996, A Coding System for Appraising Questionnaires. (In: Schwarz, N., and S. Sudman (eds.), Answering Questions, pp. 259-291.)

Martin, E., and C. Tucker, 1999, Towards a Research Agenda: Future Development and Applications of Cognitive Sciences to Surveys. (In: Sirken et al. (eds.), Cognition and Survey Research, pp. 363-381.)

Nuyts, K., H. Waege, G. Loosveldt, and J. Billiet, 1997, The Use of Cognitive Interviewing Techniques to test Measuring Instruments for Survey Research. (In Dutch: Het gebruik van cognitieve interviewtechnieken bij het testen van meetinstrumenten voor survey onderzoek. Tijdschrift voor Sociologie, No. 4, pp. 477-500.)

Oppenheim, A.N., 1992, Questionnaire Design, Interviewing and Attitude Measurement. New Edition. (Pinter Publishers, London.)

Pierzchala, M., and T. Manners, 1998, Producing CAI Instruments for a Program of Surveys. (In: Couper et al. (eds.), Computer Assisted Survey Information Collection, pp. 105-123.)

Rothgeb, J.M., L.S. Loomis, and J.C. Hess, 2000, Challenges and Strategies in gaining Acceptance of Research Results from Cognitive Questionnaire Testing. (Paper presented at the Fifth International Conference on Social Science Methodology, October 3-6, Cologne, Germany.)

Saris, W.E., 1998, The split-ballot MTMM experiment: An alternative way to evaluate the quality of questions. (Research paper, University of Amsterdam, Amsterdam.)

Schwarz, N., 1999, Cognitive Research into Survey Measurement: Its Influence on Survey Methodology and Cognitive Theory. (In: Sirken et al. (eds.), Cognition and Survey Research, pp. 65-75.)

Schwarz, N., and S. Sudman (eds.), 1996, Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. (Jossey-Bass, San Francisco.)

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds.), 1999, Cognition and Survey Research. (Wiley, New York.)

Snijkers, G. 1992, Computer-Assisted Interviewing: Telephone or Personal ? A Literature Study. (In: Westlake et al. (eds.), Survey and Statistical Computing, pp. 137-146.)

Snijkers, G., 1995a, What is the total net household income? Cognitive interviews on income questions. (Paper presented at the International Conference on Survey Measurement and process Quality, April 1-4, Bristol, UK.)

Snijkers, G., 1995b, Pre-tests on ECHP Questions on Daily Activities, Pension Schemes, and Education and Training: Final Results. (Internal CBS report, BPA no. H8170-95-GWM. Statistics Netherlands, Department of Data Collection Methodology, Heerlen.)

Snijkers, G., 1997, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. (Bulletin de Methodologie Sociologique, No. 55, pp. 93-107.)

Snijkers, G., E. de Leeuw, D. Hoezen, and I. Kuijpers, 1999, Computer-Assisted Qualitative Interviewing: Testing and Quality Assessment of CAPI and CATI Questionnaires in the field. (In: Banks et al. (eds.), … Leading Survey and Statistical Computing into the New Millennium, pp. 231-258.)

Snijkers, G., and M. Luppes, 2000, The Best of Two Worlds: Total Design Method and New Kontiv Design. An operational Model to improve Respondent Co-operation. (Paper presented at the Second International Conference on Establishment Surveys (ICES-II): Survey Methods for Businesses, Farms, and Institutions, June 17-21, Buffalo, New York.) (In: American Statistical Association, Proceedings of Invited Papers, pp. 361-371. Alexandria, Virginia.)

Tanur, J.M. (ed.), 1992, Questions about Questions: Inquiries into the Cognitive Bases of Surveys. (Russel Sage Foundation, New York.)

Tourangeau, R., and K.A. Rasinski, 1988, Cognitive Processes underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol. 103, No. 3, pp. 299-314.)

Tucker, C., 1997, Methodological Issues surrounding the Application of Cognitive Psychology in Survey Research. (Bulletin de Methodologie Sociologique, No. 55, pp. 67-92.)

USCB (U.S. Census Bureau), 1998, Pretesting Policy and Options: Demographic Surveys at the Census Bureau. (U.S. Department of Commerce, Washington, DC.)

Van Berkel, K., J. van den Brakel, H. Lautenbach, A. Luiten, J. Michiels, J. Schiepers, G. Snijkers, J. de Ree, J. van der Valk, M. Vosmer, 2001, Quality Demands to CAPI Interviewing. (In Dutch: Kwaliteitseisen voor enquêtering via CAPI. CBS report. Statistics Netherlands, Division of Social and Spatial Statistics, Department of Development and Support, Heerlen.)

Van der Zouwen, H., and W. Dijkstra, 1998, The Interview reviewed: What does the Course of Interactions in Survey Interviews tell us about the Quality of Question Wordings? (In Dutch: Het vraaggesprek onderzocht: Wat zegt het verloop van de interactie in survey-interviews over de kwaliteit van de vraagformulering. Sociologische Gids. Vol. 45.)

Westlake, A., R. Banks, C. Payne, and T. Orchard (eds.), 1992, Survey and Statistical Computing. (North-Holland, Elsevier Science Publishers, Amsterdam.)

Willis, G.B., Th.J. DeMaio, and B. Harris-Kojetin, 1999, Is the Bandwagon headed to the Methodological Promised Land? Evaluating the Validity of Cognitive interviewing Techniques. (In: Sirken et al. (eds.), Cognition and Survey Research, pp. 133-153.)

Willis, G.B., and S. Schechter, 1997, Evaluation of Cognitive Interviewing Techniques: Do the Results generalize to the Field? (Bulletin de Methodologie Sociologique, No. 55, pp. 40-66.)

Appendix 10.1. The 'TAP' paradigm for constructing questions

(Foddy, 1993, p. 193)

The key principles explicated in this text can be summarised under the acronym: '**TAP**'. Since 'Tapping' valid, reliable, respondent information is the primary aim underlying the use of questions in social research, the 'TAP' acronym is a useful reminder of the three issues that researchers should keep in mind when they are constructing questions for interviews and questionnaires.

Topic

The topic should be properly defined so that each respondent clearly understands what is being talked about.

Applicability

The applicability of the question to each respondent should be established: respondents should not be asked to give information that they do not have.

Perspective

The perspective that respondents should adopt, when answering the question, should be specified so that each respondent gives the same kind of answer.

Appendix 10.2. Principles of good question design

(Fowler, 1995, p. 103)

Principle 1: The strength of survey research is asking people about their firsthand experiences: what they have done, their current situations, their feelings and perceptions.

- *Principle 1a*: Beware of asking about information that is only acquired *secondhand*.
- *Principle 1b*: Beware of hypothetical questions.
- *Principle 1c*: Beware of asking about causality.
- *Principle 1d*: Beware of asking respondents about solutions to complex problems.

Principle 2: Ask one question at a time.

- *Principle 2a*: Avoid asking two questions at once.
- *Principle 2b*: Avoid questions that impose unwarranted assumptions.
- *Principle 2c*: Beware of questions that include hidden contingencies.

Principle 3: A survey question should be worded so that every respondent is answering the same question.

- *Principle 3a*: To the extent possible, the words in questions should be chosen so that all respondents understand their meaning and all respondents have the same sense of what the meaning is.
- *Principle 3b*: To the extent that words or terms must be used that have meanings that are likely not to be shared, definitions should be provided to all respondents.
- *Principle 3c*: The time period referred to by a question should be nambiguous.
- *Principle 3d*: If what is to be covered is too complex to be included in a single question, ask multiple questions.

Principle 4: If a survey is to be interviewer administered, wording of the questions must constitute a complete and adequate script such that, when interviewers read the question as worded, respondents will be fully prepared to answer the question.

- *Principle 4a*: If definitions are to given, they should be given before the question itself is asked.
- *Principle 4b*: A question should end with the question itself. If there are response alternatives, they should constitute the final part of the question.

Principle 5: Clearly communicate to all respondents the kind of answer that constitutes an adequate answer to a question.

- *Principle 5a*: Specify the number of responses to be given to questions for which more than one answer is possible.

Principle 6: Design survey instruments to make the task of reading questions, following instructions, and recording answers as easy as possible for interviewers and respondents.

Principle 7: Measurement will be better to the extent that people answering questions are oriented to the task in a consistent way.

Appendix 10.3. Key decision guide: Question utility

(Czaja & Blair, 1996, p. 61)

A.	Does the survey question measure some aspect of one of the research questions?
B.	Does the question provide information needed in conjunction with some other variable? {IF NOT TO BOTH A AND B, DROP THE QUESTION. IF YES TO ONE OR BOTH, PROCEED.}
C.	Will most respondents understand the question and in the same way? {IF NO, REVISE OR DROP. IF YES, PROCEED.}
D.	Will most respondents have the information to answer it? {IF NO, DROP. IF YES, PROCEED.}
E.	Will most respondents be willing to answer it? {IF NO, DROP. IF YES, PROCEED.}
F.	Is other information needed to analyze this question? {IF NO, PROCEED. IF YES, PROCEED IF THE OTHER INFORMATION IS AVAILABLE OR CAN BE GOTTEN FROM THE SURVEY.}
G.	Should this question be asked of all respondents or of a subset? {IF ALL, PROCEED. IF A SUBSET, PROCEED ONLY IF THE SUBSET IS IDENTIFIABLE BEFOREHAND OR THROUGH QUESTIONS IN THE INTERVIEW.}

Samenvatting

In de literatuur over vragenlijstontwerp en survey methodologie wordt pre-testen genoemd als een manier om vragenlijsten te evalueren (te onderzoeken of ze werken zoals bedoeld) en om het optreden van meetfouten te kunnen vaststellen (dat wil zeggen het vaststellen van de datakwaliteit, in het bijzonder de validiteit). De *American Statistical Association* (ASA, 1999, p. 11) zegt het als volgt: "De vragenlijstontwerper moet de noodzaak inzien om te **pre-testen**, te **pre-testen** en nog eens te **pre-testen**." Clark en Schober (1992, p. 29) geven aan waarom pre-testen noodzakelijk is: "Het is voor onderzoekers onmogelijk om perfecte vragen op te stellen; vragen, die voor elke respondent direct duidelijk zijn en die geen verduidelijking behoeven. En omdat ze dit niet kunnen, zullen de antwoorden vaak verrassend zijn."

In de alledaagse praktijk van het ontwerpen van een steekproef- (of survey) onderzoek, worden pre-testonderzoek en de resultaten hiervan, echter, niet altijd door die onderzoekers geaccepteerd. Een algemene houding ten aanzien van het pre-testonderzoek is verwoord door Converse en Presser (1986, pp. 51-52): "Alhoewel geadviseerd wordt om een vragenlijst altijd te pre-testen -geen handboek in de survey methodologie zal dit wetenschappelijke advies tegenspreken-, wordt hiertegen in de praktijk vaak gezondigd. Er is altijd gebrek aan geld of, als de deadlines in zicht komen, gebrek aan tijd om voldoende te pre-testen. Het gevolg is dat de pre-testen intuïtief en informeel worden uitgevoerd. Zo zijn er geen algemene regels die aangeven hoe een goed pre-testonderzoek moet worden uitgevoerd, er is geen systematisering van de praktijk, er bestaat geen overeenstemming over wat van pre-testonderzoek wordt verwachten, en er wordt slechts zelden verslag gedaan over de uitvoering. Hoe pre-testonderzoek wordt uitgevoerd, wat de onderzoekers ervan hebben geleerd, en hoe ze vervolgens de vragenlijst hebben aangepast – er wordt slechts summier over deze aspecten gerapporteerd in onderzoeksrapporten, als er al gerapporteerd wordt. Het is daarom niet verbazingwekkend dat de invloed van pre-testonderzoek soms wordt overdreven, terwijl de mogelijkheden vaak worden onderschat."

In dit proefschrift heb ik geprobeerd mijn ervaringen met pre-testonderzoek bij het Vragenlab van het Centraal Bureau voor de Statistiek, een cognitief laboratorium dat in 1992 van start ging, te beschrijven. Het doel van dit proefschrift is documentatie van deze praktijkervaringen. Deze dissertatie bestaat niet uit een theoretische verhandeling over cognitieve laboratorium methodes, maar beschrijft de toepassing van deze methodes: het opzetten en uitvoeren van pre-testonderzoek, het analyseren van de gegevens en het presenteren van de resultaten. Voor zover bij mij bekend is er in de literatuur over Cognitieve Aspecten van Survey Methodologie (CASM) weinig gerapporteerd over hoe dit soort onderzoek in de praktijk wordt uitgevoerd. Dit wordt bevestigd door Willis e.a. (1999, p. 137), die schema's voor het systematisch beschrijven van de uitvoering van cognitieve interview methodes in de praktijk bediscussiëren. Zij concluderen dat "(…) zo'n schema niet bestaat voor het gebruik van cognitieve interviews." In dit proefschrift heb ik getracht een schematische en systematische beschrijving te geven van de uitvoeringspraktijk van

pre-testonderzoek bij het Vragenlab van het CBS, om zo als uitgangspunt te kunnen dienen voor andere onderzoekers. Ik hoop dat ik hierin geslaagd ben.

Cognitief (pre-)testonderzoek is gericht op verbetering van de kwaliteit van onderzoeksgegevens, door verbetering van de vragenlijst. De vragenlijst wordt gevalideerd door deze op kleine schaal uit te testen, dat wil zeggen dat fouten in de vragenlijst -die systematische fouten veroorzaken in het vraag-en-antwoord proces bij de respondent in een interview- worden opgespoord, verklaard en verbeterd (in een iteratief proces). Op deze manier wordt de vragenlijst aangepast aan het vraag-en-antwoord proces, wordt deze gemakkelijker te beantwoorden, kost dat minder tijd en wordt deze respondentvriendelijker, met als resultaat vermindering van meetfouten, dat wil zeggen verhoging van de datakwaliteit (de interne validiteit), en verlaging van de respondentbelasting. Dit is het CASM-paradigma. Het CASM-paradigma definieert de centrale hypothese van dit proefschrift: het pre-testen in een cognitief laboratorium valideert vragenlijsten.

Het vraag-en-antwoord proces is beschreven door Tourangeau en Rasinski. In 1988 presenteerden zij een 4-stappen model. Volgens dit model bestaat het vraag-en-antwoord proces uit de volgende 4 stappen: interpretatie van de vraag, opdiepen van de benodigde informatie om tot een antwoord te komen, evaluatie en samenvoeging van deze informatie tot een antwoord, en rapportage van dat antwoord. Door toepassing van cognitief onderzoek op informatieverwerkings-processen die spelen tijdens het beantwoorden van enquêtevragen (Jobe & Mingay, 1991, p. 178) “werd de respondent beschouwd als een vraag-en-antwoord systeem dat een aantal mentale operaties uitvoert, zoals het begrijpen van wat met de vraag wordt bedoeld, het opdiepen van relevante informatie uit het geheugen, besluitvorming om tot een antwoord te komen en het vermelden van dat antwoord aan de interviewer.” Volgens Jobe en Mingay (1991, p. 178) “betekende het modelleren van het denkproces van de respondent een grote verandering ten opzicht van het eenvoudige stimulus-respons model van responsgedrag, dat vanaf het begin van het moderne survey onderzoek het uitgangspunt vormde voor het ontwerpen van meetinstrumenten.”

Het vraag-en-antwoord proces wordt onderzocht in een cognitief laboratorium. Pre-testonderzoek, zoals uitgevoerd in zo’n laboratorium, bestaat in de regel uit een beperkt aantal cognitieve interviews waarin respondenten worden gevraagd om informatie over hun mentale proces te verschaffen terwijl ze vragen beantwoorden. Methodes die bij het Vragenlab worden gebruikt om het vraag-en-antwoord proces te onderzoeken, zijn: de expert beoordeling, de focusgroep discussie, het diepte-interview (waarin gebruik wordt gemaakt van verschillende technieken zoals hardop denken, open doorvragen, gericht doorvragen naar de betekenis van begrippen, parafraseren, gerichte testvragen, en vignettes), en de gedragscodering.

Dit proefschrift bestaat voor een groot deel uit rapporten die reeds eerder zijn gepubliceerd als CBS-rapport, onderzoeksrapport of als een bijdrage aan een internationaal congres. In deze rapporten worden onderzoeksprojecten beschreven die zijn uitgevoerd in de periode 1992-2000. Door te kiezen voor een samenstelling van het proefschrift op basis van afzonderlijke rapporten,

bestaat er hier en daar enige overlap tussen de hoofdstukken. Om van het proefschrift toch één geheel te maken, zijn een aantal rapporten marginaal bewerkt. Dit is op een zodanige wijze gedaan dat elk hoofdstuk alsnog afzonderlijk gelezen kan worden.

Het eerste hoofdstuk is een inleidend hoofdstuk, waarin cognitief laboratorium onderzoek wordt geïntroduceerd. In dit hoofdstuk worden ook de geschiedenis van de CASM-beweging (die haar oorsprong heeft in de Verenigde Staten in de jaren ’80 van de vorige eeuw) en de geschiedenis van het Vragenlab van het CBS beschreven. Tevens wordt een overzicht van het proefschrift gegeven. De volgende twee hoofdstukken gaan over computergestuurd interviewen. De meeste vragenlijsten van het CBS worden in een computergestuurd interview afgenomen. Hiermee worden de condities voor het pre-testen in het Vragenlab bepaald. In hoofdstuk 2 komt het computergestuurde interview aan de orde; in hoofdstuk 3 worden de effecten van computergestuurd interviewen op de datakwaliteit beschreven. Deze drie hoofdstukken vormen de inleiding voor de andere hoofdstukken, de eigenlijke kern van het proefschrift. In deze hoofdstukken wordt de toepassing van vragenlab methodes beschreven, alsmede enkele case studies.

De methodes die in het Vragenlab van het CBS worden gebruikt, zijn het onderwerp van de hoofdstukken 4 en 5. Een overzicht van die methodes wordt gegeven in hoofdstuk 4. Het betreft: de expert beoordeling, de focusgroep discussie, het diepte-interview (waarin gebruik wordt gemaakt van verschillende technieken zoals hardop denken, open doorvragen, gericht doorvragen naar de betekenis van begrippen, parafraseren, gerichte testvragen, en vignettes), en de gedragscodering. Deze methodes worden op basis van praktijkervaring beschreven, dat wil zeggen dat beschreven wordt hoe ze zijn toegepast in het Vragenlab. Computergestuurd kwalitatief interviewen (CAQI) is het onderwerp van hoofdstuk 5. Deze methode is bij het Vragenlab van het CBS ontwikkeld om computergestuurde vragenlijsten te kunnen pre-testen. Met CAQI wordt het pre-test protocol geïntegreerd in de te testen computergestuurde vragenlijst.

In de volgende vier hoofdstukken wordt cognitief onderzoek aan de hand van een aantal case studies geïllustreerd. In deze hoofdstukken komen het ontwerp van deze studies en de resultaten hiervan (zoals geïdentificeerde problemen in de vragenlijst en aanbevelingen voor verbetering) aan de orde. De hoofdstukken 6 en 7 bestaan uit onderzoeksrapporten, waarmee deze hoofdstukken tevens als illustratie dienen van hoe de desbetreffende onderzoeken en de resultaten aan de klant zijn gepresenteerd.

In hoofdstuk 6 en 7 komen vragenlabonderzoeken waarin CAQI diepte-interviews zijn toegepast, aan de orde. Hoofdstuk 6 handelt over het pre-testen van inkomensvragen. Deze vragen zijn onderzocht met behulp van hardop denken met open doorvragen, gericht doorvragen en gerichte testvragen. In hoofdstuk 7 worden vragen uit het Europese Huishoudens Panel (European Community Household Panel, ECHP) onderzocht. Het betreft vragen over dagelijkse activiteiten, pensioenen, en opleiding en training. In dit onderzoek zijn hardop denken met open en gericht doorvragen, en vignettes toegepast.

Hoofdstuk 8 handelt over een CAQI veldonderzoek waarin de vragenlijst van het Permanent Onderzoek Leefsituatie (POLS) is onderzocht. In deze kwalitatieve operationele veldtest zijn respondent gedragscodering en gerichte testvragen toegepast. (In het vragenlabonderzoek dat aan de veldtest vooraf ging, zijn diepte-interviews en focusgroep discussies gehouden.)

In hoofdstuk 9 wordt een onderzoek met focusgroep discussies beschreven. Met dit onderzoek is een nieuw formulier voor de jaarlijkse Productiestatistiek, een bedrijfsonderzoek, getest. Terwijl in de voorgaande case studies vragenlabonderzoek is gebruikt om vragenlijsten voor persoons- en huishoudensenquêtes te testen, wordt in dit hoofdstuk vragenlabonderzoek toegepast om een formulier van een bedrijfsenquête te testen. Dit onderzoek wordt in dit hoofdstuk gepresenteerd in een discussie over actieve berichtgeverscommunicatie. Actieve berichtgeverscommunicatie is gericht op het motiveren en stimuleren van mensen die voor een onderzoek worden benaderd, om mee te doen aan dat onderzoek. De rol van een vragenlab in deze strategie is het ontwikkelen van goede formulieren. Dat zijn formulieren die er aantrekkelijk uitzien, en gemakkelijk zijn te begrijpen en in te vullen, om zo de respondentbelasting te verminderen. In dit hoofdstuk wordt cognitief laboratorium onderzoek geplaatst in een breder kader van *Total/Tailored Survey Design* en *Total Quality Management*.

Hoofdstuk 10 besluit dit proefschrift met de conclusies. In dit hoofdstuk wordt een samenvatting van het proefschrift gegeven, waaronder een samenvatting van de besproken vragenlab methodes en van de case studies. Een aantal van de gesignaleerde problemen in de onderzochte vragenlijsten betreft: het gebruik van technische (jargon) of onduidelijke woorden, complexe zinsbouw, lange vraag, dubbele vraagstelling, volgorde effect op basis van voorafgaande vraag, er wordt gevraagd naar specifieke informatie die niet direct uit het geheugen kan worden opgediept, er moet een complexe berekening worden uitgevoerd om tot het antwoord te komen, of er zijn overlappende of ontbrekende antwoordcategorieën. In dit hoofdstuk worden de gesignaleerde problemen gerelateerd aan ontwerpfouten in de vragen op basis van ontwerpprincipes voor vragen.

Alhoewel pre-test onderzoek een manier is om vragenlijsten te evalueren en meetfouten te reduceren, worden de resultaten van dit soort onderzoek -zoals aanbevelingen voor verbering van de vragenlijst- niet altijd geaccepteerd. In dit laatste hoofdstuk wordt tevens de kritiek op dit soort onderzoek besproken, maar ook worden strategieën genoemd om dit soort onderzoek en de resultaten ervan te presenteren en om de acceptatiegeneigdheid te vergroten. Eén van deze strategieën is uitvoering en documentatie van het onderzoek volgens wetenschappelijke criteria. Ter completering van dit proefschrift worden ook enkele specifieke testen gericht op het verbeteren van het computeriseren van vragenlijsten besproken. Het betreft testen waarmee het programmeringsproces van gecomputeriseerde meetinstrumenten en de mens-machine interactie worden onderzocht. Dit hoofdstuk besluit met enkele ideeën voor verder onderzoek, en tot slot wordt geconcludeerd dat het pre-testen in een cognitief laboratorium vragenlijsten valideert.

Referenties

ASA, 1999, Designing a Questionnaire. ASA series: What is a Survey? (American Statistical Association, Alexandria, VA.)

Clark, H., and M. Schober, 1992, Asking Questions and Influencing Answers. (In: Tanur (ed.), Questions about Questions, pp. 15-48.)

Converse, J.M., and S. Presser, 1986, Survey Questions. Handcrafting the Standardized Questionnaire. (Quantitative Applications in the Social Sciences Series, No. 63, Sage, Beverly Hills.)

Jobe, J.B., and D.J. Mingay, 1991, Cognition and Survey Measurement: History and Overview. (Applied Cognitive Psychology, Special Issue on Cognition and Survey Measurement, Vol. 5, No. 3, pp. 175-192.)

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds.), 1999, Cognition and Survey Research. (Wiley, New York.)

Tanur, J.M. (ed.), 1992, Questions about Questions: Inquiries into the Cognitive Bases of Surveys. (Russel Sage Foundation, New York.)

Tourangeau, R., and K.A. Rasinski, 1988, Cognitive Processes underlying Context Effects in Attitude Measurement. (Psychological Bulletin, Vol. 103, No. 3, pp. 299-314.)

Willis, G.B., Th.J. DeMaio, and B. Harris-Kojetin, 1999, Is the Bandwagon headed to the Methodological Promised Land? Evaluating the Validity of Cognitive interviewing Techniques. (In: Sirken et al. (eds.), Cognition and Survey Research, pp. 133-153.)

Curriculum vitae

7 February 1963	Ger Snijkers born in Schaesberg (Landgraaf), The Netherlands	
1975 - 1981	Secondary school: Atheneum B, Eijkhagen College, Landgraaf	
1981 - 1987	Doctoral Political Science (cum laude), with specialisation in social science methodology, University of Amsterdam	
1987 - 1988	Propaedeutics Mathematics, Free University, Amsterdam	
1987 -	Employee of Statistics Netherlands (CBS), in several appointments:	
	1987 - 1992	Junior researcher at the Department of Statistical Methods (CBS-Voorburg)
	1992 - 1995	Researcher at the Questionnaire Laboratory (Department of Data Collection Methodology, CBS-Heerlen)
	1995 - 2000	Senior researcher and project leader of research and advisory projects at the Questionnaire Laboratory (Department of Data Collection Methodology, CBS-Heerlen)
	2000 -	Manager of research and development projects with regard to survey methodology and non-response at the Department of Development and Support (Division of Social and Spatial Statistics, CBS-Heerlen)