Questionnaire development

12

Deidre Giesen, Vivian Meertens, Rachel Vis-Visschers and Dirkjan Beukenhorst

Statistical Methods (201210)

Statistics Netherlands

The Hague/Heerlen, 2012

Explanation of symbols

.	data not available
*	provisional figure
**	revised provisional figure (but not definite)
x	publication prohibited (confidential figure)
–	nil
–	(between two figures) inclusive
0 (0.0)	less than half of unit concerned
empty cell	not applicable
2011–2012	2011 to 2012 inclusive
2011/2012	average for 2011 up to and including 2012
2011/’12	crop year, financial year, school year etc. beginning in 2011 and ending in 2012
2009/’10–2011/’12	crop year, financial year, etc. 2009/’10 to 2011/’12 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands
Grafimedia

Cover
Teldesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form:
www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1876-0333

60165201210 X-37

Table of contents

1. Introduction to the theme

1.1 General description and reading guide

This part of the Methods Series deals with questionnaire development and evaluation, which for the sake of brevity is referred to hereafter as questionnaire development. Questionnaires may be developed for new statistical products, and when evaluating or redesigning existing products and processes. Questionnaire development as discussed in this document is applicable to all survey modes (CAPI, CATI, web questionnaires and paper questionnaires) commonly used within Statistics Netherlands and for all target populations (questionnaires for individuals and households and for businesses and organizations).

Questionnaire development has three, sometimes conflicting, objectives. The first is that a questionnaire must facilitate the collection of the data needed to answer the research questions, which requires the data to be valid, reliable and relevant (see Chapter 3). Then the survey must be as user-friendly and convenient as possible for the respondent and interviewer. Finally it must be possible to develop, process and modify questionnaires efficiently. It may be important to be able to respond rapidly to events and changing circumstances, not only for new themes, but also where revision is needed because of changes in legislation (e.g. the reform of health insurance).

This part of the Methods Series is based on the extensive literature, experience within Statistics Netherlands, and the 'Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System' (Brancato et al., 2006). The guidelines given in this handbook correspond to the requirements set down in the European Statistics Code of Practice for the National and Community Statistical Authorities (Eurostat, 2012). One of the quality indicators is: 'In the case of statistical surveys, questionnaires are systematically tested prior to the data collection.'.

This part of the Methods Series has the following structure. Chapter 2 describes the entire questionnaire development process. Chapter 3 then comprehensively covers the questionnaire design part. We set out several central theoretical principles in questionnaire design, such as the question-answer process, leading to firm guidelines for phrasing survey questions and structuring and formatting questionnaires. Chapter 4 discusses the methods and techniques used in testing and evaluating questionnaires in some depth. We cover the methods traditionally used in a laboratory setting, such as cognitive interviews, and methods applicable to questionnaires already in the field, such as evaluating questionnaires with interviewers.

This part of the Methods Series concludes with various appendices with detailed information, including a questionnaire design checklist.

1.2 Scope and relationship with other themes

This theme of the Methods Series describes the questionnaire development process, with descriptions of the methods involved, including various test procedures. These methods are an essential part of questionnaire development.

There is no coverage of specific operational processes within Statistics Netherlands or detailed technical descriptions, which are outside the scope of the Methods Series.

A new questionnaire development process starts only if the information needed cannot be obtained from existing surveys or registers. This rule is in line with the Statistics Netherlands data collection policy of collecting primary data only when secondary information is unavailable (Beukenhorst, Van Dam, Göttgens, Snijkers & Verbruggen, 2005). Other themes and subthemes of the Methods Series, e.g. Secondary Data Collection and XBRL, cover matters concerned with accessing data held in registers, etc. Other related Methods Series themes are Organization of Fieldwork and Data collection Strategy (including Mixed Mode Approach).

1.3 Place in the statistical process

In terms of statistical process as defined within the Methods Series by Willeboordse (2008), questionnaire development resides mainly in the design phase. Questionnaire development starts with specifying the required output, i.e. determining which statistics are ultimately to be delivered, and ends with evaluating the data collection process and the quality of the data actually collected. See also Chapter 2 for more extensive coverage of questionnaire development as a process.

1.4 Definitions

Term	Definition
Business respondent	Term used in business surveys: the person or organization that answers the questions. The equivalent of a respondent in a population survey.
CAPI	Computer Assisted Personal Interviewing, also known as face-to-face interviewing. Interviews in which the interviewer is present in person and uses a computer-assisted questionnaire with the survey questions. CAPI surveys are usually administered at respondents' homes.
CATI	Computer Assisted Telephone Interviewing, which is surveying by telephone using a computer-assisted questionnaire.
CAWI	Computer Assisted Web Interviewing, which is a computer-assisted questionnaire in which respondents themselves answer the questions on the Internet.
Nonresponse	• Unit nonresponse: observation unit for which no questionnaire data were obtained; • Item nonresponse: question in a questionnaire to which no answer was obtained (where one or more other questions were answered, i.e. not in conjunction with unit nonresponse).
Survey target person	The person of whom data is collected. This may be, but is not necessarily, the respondent. See respondent.
PAPI	Paper and Pencil Interviewing. Paper questionnaires to be completed by respondents themselves. They may be sent by post or distributed by an interviewer.
Primary data	Within Statistics Netherlands: data collected by or on behalf of Statistics Netherlands.
Respondent	The person who answers the questions. This will usually be the survey target person, but the respondent could be acting as a proxy for this individual, or answering questions on behalf of a household.
Statistical unit	Unit about which the statistic reports.
Secondary data	Within Statistics Netherlands: data collected by parties other than Statistics Netherlands (but not on behalf of Statistics Netherlands).
Questionnaire	The set of questions put to a respondent.
Questionnaire design	The process of creating the questionnaire texts, structure and format.
Questionnaire development	The entire process involved in drawing up a questionnaire: from conceptualization and research design to final evaluation.
Observation unit	The unit about which data are being collected. Observation units may later be aggregated into statistical units, or disaggregated.
Survey	This term has two meanings: • A form of research in which selected people or other units are investigated in the standard manner; •a synonym for the standard questionnaire itself.

2. Questionnaire development as a process

Questionnaire development involves more than drawing up and formatting question texts. Questionnaire development is a process that may consist of the six steps set out below (see Figure 2.1, based on Brancato et al., 2006):

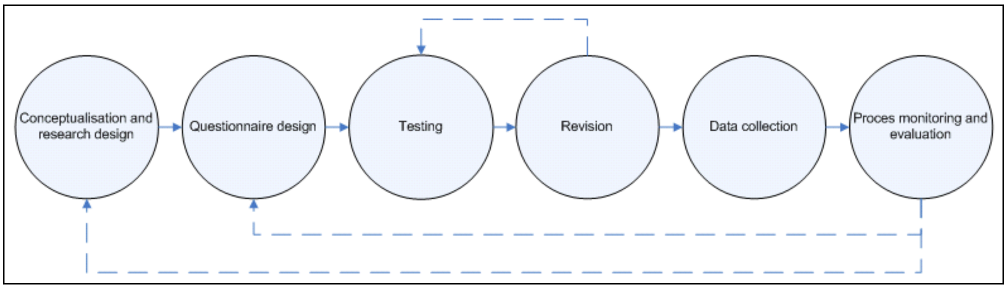

Figure 2.1 Questionnaire development as a process

The six steps and the arrows joining them show that the questionnaire development process is iterative, in that information from each phase in the process may serve as input for a subsequent or previous phase. Besides questionnaire specialists, this process also involves content matter specialists and other experts.

2.1 Conceptualization and research design

The first step in questionnaire development is conceptualization, which involves defining the subject and the variables to be measured. Theoretical survey concepts are often complex, such as poverty, access to healthcare, outsourcing, or being a victim of crime. It is impossible to translate concepts of this kind into survey questions on a one-to-one basis. It must first be established exactly what the clients or users of the data want to know, and then to assess the feasibility of measuring these concepts in the target population. This process must ultimately lead to concrete, quantifiable concepts, which are referred to as indicators. Content matter experts will be closely involved in defining the indicators. The indicators must then be translated into one or more survey questions. This final part of the process is known as operationalization, in which the role of the questionnaire developers comes more to the fore. For example, if the concept is poverty, it will be necessary to define precisely what poverty means. A possible operationalization is: a single person can be deemed poor if their income is less than 50% of the minimum wage, or a family can be deemed poor if its income is less than 70% of the minimum wage. The questionnaire would accordingly include questions about respondents' income and household composition. For some considerable time Statistics Netherlands has used the Tax Authority as a source of personal and household income data.

Satisfactory operationalization is served by consulting existing professional literature, internal documents and experts (by reviewing existing documents or conducting an expert review). Besides professional literature about content matter and methods, other examples would be internal documents with reports of interviewer evaluations or quality analyses of relevant pre-existing questionnaires, reports from statistical institutes in other countries, and experts with knowledge of the subject or field under investigation. Try to track down the questionnaires that were used in any previous internal or external studies of the subject, and find out how well they functioned.

Where entirely new subjects or target populations are involved, it makes sense in this phase to gauge how well potential respondents understand the envisaged research design and the concepts to be measured. This can be achieved with structured group discussions (focus groups) or in-depth interviews with test respondents (cognitive interviews). Both methods are covered in Chapter 4.

Having established what is to be measured, it must be decided who is to be subjected to the measurement, and how it is to be performed. The first thing is to determine the relevant statistical unit. Does the information being sought relate to a person, or to a household, or a business unit, or a department within a company? The statistical unit is the unit on which the statistic is based. Another thing to establish is the level on which the data is to be observed. The observation unit for business data collection may be larger or smaller than the statistical unit being reported (parent group, holding company, business location, etc). The business respondent is the party that actually completes the questionnaire. The observation unit for a personal data collection may be an individual (the survey target person) or a household. The respondent might not be the same as the survey target person (who could be a child, in which case the questions would be answered by a parent or carer). Respondents can also provide information about their entire household, or all its members. Other important choices are whether to engage an interviewer and which data collection mode to use. Experience with surveys carried out in the past may help when making these decisions, and an examination of response lists, statistical analyses and other outputs can certainly prove useful. It may be useful to venture into the field when choosing a research design. For instance, for a new survey into the outsourcing of business operations, potential respondents were asked who in their company would be best placed to complete the questionnaire (Vis-Visschers, 2007).

In summary, the phase of conceptualization and research design therefore defines the survey terrain. Conceptualization involves specifying what is to be measured with the questionnaire, or, in other words, which variables the questionnaire must deliver. The research design gives rise to a set of conditions that the questionnaire must satisfy.

Finally: document all choices!

Surveys of National Statistical Institutes often remain in the field for some considerable time, or are repeated after a long intervening period. Furthermore, people tend to move around within an organization, which means that knowledge about the development process may disappear. Thorough documentation of the choices and considerations in this phase is therefore of great importance.

Documentation of this kind is valuable both to later users of the data and to questionnaire developers when they review the survey.

2.2 Questionnaire design

A textual version of the questionnaire is compiled and formatted in the questionnaire design and construction phase. The survey population and the research design, and in particular the data collection mode, are clearly important constraints. For example, a paper questionnaire provides far fewer options for complex routing than a computer counterpart. For additional information about choosing the data collection mode see the Data Collection Strategy theme of the Methods Series.

Questionnaire design is ideally a team activity, where the team members are experts in the following areas: the content of the survey (usually the statistical department), questionnaire design and the logistics of data collection and processing. It is also advisable with complex and new PAPI and CAWI questionnaires to solicit the advice of a graphical designer.

One of the first steps in this phase is to devise a questionnaire structure that respondents will find logical. This step sets down the sequence in which the various topics are presented (and, if applicable, the order in which respondents are to become involved). The question sequence for each topic is then determined. In the process, the variables defined in the previous phase are translated into specific survey questions. When choosing the variables and questions it must be borne in mind which question is relevant for which respondent. Routing is used to lead respondents to the questions that are relevant for them.

A questionnaire is formatted once a version of the routing and the texts is available, since texts are presented in a certain format depending on the mode of data collection and the standards applied in the organization. Points to be considered include interviewer instructions, explanations to question texts, headings above interrelated blocks of questions and information about the routing. For PAPI and CAWI questionnaires in particular, where the respondents themselves see the questionnaire, formatting is an essential part of the design phase, and should start as early as possible. For example, it can sometimes be necessary to adapt texts or question sequences to fit the space available on a page or display screen.

When modifying an existing questionnaire, there must be a review of any available documents about monitoring and evaluations with interviewers or respondents of the already implemented survey. The information gained may lead to changes in questionnaire design or the questionnaire itself. Reports of focus groups with interviewers and respondents serve the same goal. Questionnaire design is covered in depth in Chapter 3, and focus groups are described in Chapter 4.

2.3 Testing

Testing a draft questionnaire can serve a variety of objectives. For example, a computer-assisted questionnaire could be tested for correct construction, or for how well respondents comprehend and correctly answer crucial questions, or to confirm that the questionnaire is user-friendly for the interviewer and respondent, and ascertain how long it will take to complete. Eurostat requires all new questionnaires to be tested at least once with potential respondents (Brancato et al., 2006, page XI). For questionnaires that are already in the field, or when a questionnaire is to be revised, this requirement entails an analysis of the costs and benefits involved. The most commonly used methods in this phase of questionnaire development are presented briefly below. Additional details of the most important of these methods are given in Chapter 4.

Informal tests

Informal tests are the simplest to perform. For example, questionnaire developers might complete their own questionnaire using real answers, read all the texts out loud, or interview someone they know. Testing in this way gives an initial impression of how easy the questions are to read out, understand and answer.

Expert review

In an expert review a group of experts assesses a questionnaire.

Small-scale, qualitative tests

The test phase often involves small-scale, qualitative tests to assess the respondent and interviewer friendliness of certain questions, or the entire questionnaire. This might be achieved through *focus groups*, *cognitive interviews* with test subjects, or both. *Observations* of test respondents completing a questionnaire, or interviewers administering a questionnaire, are also valuable at this stage, and can be used in combination with the methods mentioned earlier.

Functional or technical testing

It is important with computer-assisted questionnaires to test thoroughly whether the questionnaire structure complies with the specification and runs in the correct sequence. All the routings must be checked for correctness and it must be confirmed that specific groups are presented with all the questions relevant for them. It is also important to ascertain whether the correct follow-up question is given after the answer options ‘don’t know’ or ‘refusal’. The functional or technical test is needed to achieve this aim.

Small-scale, operational tests

An operational test, which may also be referred to as a pilot or dress rehearsal, involves trying out the complete data collection procedure on a small scale. A focus of testing is on how well the logistic procedures work in the field.

Large-scale pilots and trials

The data collected in large-scale pilots and trials may also be analysed quantitatively. Attention may also be given to non-responses to questions and questionnaire components and to the quality of the collected data. Experiments are useful where help is needed to decide between alternative phrasings of questions, or alternative methods, and when comparing two or more questionnaires.

2.4 Revision

A wide variety of change proposals generally emerge from testing a questionnaire and, where relevant, the data collection procedure. Whether and how the proposals are to be implemented sometimes demands extensive discussion with all concerned. It will often be necessary to strike a tricky balance between different interests, such as the need for comparability with other questionnaires or previous data collections (time series!) versus how convenient respondents and interviewers will find a questionnaire in use. The revised questionnaire should ideally be retested for content, and certainly for technical aspects, because of the possibility of new problems being introduced along with changes in the questionnaire.

Proper development and testing places heavy demands on time and money, both of which tend to be scarce in practice. Accordingly, both questionnaire development and testing must be planned with great care. There is little point in embarking on the testing of a hastily compiled questionnaire in the certain knowledge that many problems will be revealed. However, there is equally little point in spending so much time on questionnaire development that none is left for testing. If a questionnaire is used without having been tested first and no problems emerge, you can conclude that it was perhaps not so bad after all, but that does not mean that it could not have been better. Extensive testing is able to prevent many problems, but is likewise no guarantee of producing the best possible questionnaire. Proper development and testing require sufficient time to be set aside, bearing in mind that there must also be enough time to develop, test and implement any changes.

2.5 Data collection

When a questionnaire is finished it can be implemented in the field. This is the phase in which data are collected. The questionnaire may include supplementary evaluative questions for respondents or interviewers. For instance, there could be questions about the time available for the interview or for answering, how difficult the questions were, how certain respondents were about their answers to particular questions, or the interviewer's views about how well the dialogue proceeded. However, handling these questions could be inconvenient, for field interviewers in particular. This instrument must therefore be used sparingly. Methods exist that require interviewers only to indicate whether 'a problem' occurred with a given question. An investigation of how often and why any problems occurred can then be conducted at a later phase.

2.6 Process monitoring and evaluation

Questionnaire development is not over when a questionnaire has been implemented in the field. It is important with new questionnaires in particular to keep an eye on what happens during data collection. It can happen that minor adjustments are needed in spite of extensive development and testing, or that a supplementary instruction for the interviewers, help desk, respondents, or data processing has to be added during data collection to limit the consequences of the problems that occurred.

An evaluation of how a questionnaire fared in the field can yield valuable information for those who analyse the data. Furthermore, when questionnaires are reused often, as they usually are at Statistics Netherlands, field experience may prove useful in improving a questionnaire in a subsequent revision. It is then worthwhile to consider both the *process of data collection* and the output: *the collected data*.

The primary important sources of information about questionnaire performance are the respondents and the employees who come into contact with them, i.e. the interviewers, the field staff and help desk staff. Data quality can be investigated by analysing aspects such as nonresponse, how answer scales were used, the plausibility of answers, or the distribution of answers across response categories. Another useful source is completed paper forms, where crossings-out and comments are of particular interest. Yet another important source of information is respondents' questions and comments recorded by the help desk.

For the kind of long-term, repeated surveys that Statistics Netherlands performs, it is particularly important to arrange for feedback from the data collection phase to the design phase. There is additional information about the use of the above evaluation methods in Chapter 4.

3. Questionnaire design

3.1 Introduction

The texts, structure and format of a questionnaire are created in the questionnaire design phase. This phase often involves striking a difficult balance between multiple interests. For instance, it would make sense to measure the financial position of households through several questions that remind the respondent of all possible forms of wealth. However, a greater number of questions also brings additional costs (e.g. for more design time and interview time) and sometimes a greater respondent burden.[1] An excessive respondent burden can lead to careless answers and even nonresponse.

There is no one right way for striking this kind of balance, and composing a questionnaire is therefore always a creative process in which different teams will arrive at different, equally good, solutions. It is rarely possible to identify a single best solution. We list in this chapter the most important tools for creating well-thought-out questionnaires. Several general principles, such as the theoretical model of the question-answer process and the most important sources of measurement error, are presented first (Section 3.2). We then focus on the formulation of questions and answer options (Sections 3.3 and 3.4). The final aspects are how to use explanations and instructions and the structure and format of questionnaires (Sections 3.5, 3.6 and 3.7).

Much has been written about questionnaire design and this chapter therefore covers a great deal of ground. If it all appears complex, just remember the simple basic rule of questionnaire design: make sure the respondent understands what you are asking, is able to respond and is willing to do so (see Box 3.1).

Box 3.1 Basic questionnaire design rule

The respondent must:

- understand the question;
- be able to respond;
- be willing to do so.

[1] In some cases a series of related questions can actually make it easier for respondents to answer complex questions correctly.

3.2 General questionnaire design principles

3.2.1 Reliability and validity

Questionnaires are measurement instruments. Reliability and validity are measures of how well this instrument works.

Reliability indicates how closely the results of repeated measurements of the same concept agree: a reliable measurement of a quantity known not to have changed that is performed twice with the same person will produce the same value both times. The same is true where two different people are involved who do not differ on the measured characteristic. Only if this condition is met can we produce statistics from the answers to survey questions.

Validity means that the concept measured is actually the one the researcher intended. You can count apples with the utmost reliability, but if what you want is a statistic about pears, you will have failed to perform a valid measurement.

The need to perform reliable and valid measurements requires survey questions to satisfy two conditions: 1) each respondent must interpret the question in the same way; 2) each respondent must understand the question the way the researcher intended.

We use standardized surveys to satisfy these conditions: we try to ensure that the way every respondent hears or sees the survey questions is as similar as possible. For example, it would be unacceptable if some respondents but not others can see the answer options to a question because of the screen settings. The theoretical background of the standardized survey is the stimulus-response model. The theory holds that any difference in response to an identical stimulus under identical conditions will be a consequence of a real underlying difference. For example, if you suspected the presence of a hollow space somewhere behind a wall, you would rap the wall with your knuckles until you heard a hollow sound. This would work only if you always rapped the wall with your knuckles in exactly the same way. If you were suddenly to change how you hold your hand, you could not know whether the different sound you then heard was caused by a hollow space, or the different rapping technique.

The preference for standardized survey questions, as in Statistics Netherlands, makes it vital for interviewers to run through questionnaires in exactly the same way, and always to give the same explanations for questions, if necessary. The interviewer must also ask any follow-up questions that may be needed to clarify a respondent’s answer in a uniform way. Questionnaire and interviewer jointly constitute the ‘measurement instrument’ that is required to generate identical stimuli. The quality is ultimately determined by the interviewer: a good questionnaire poorly administered will not yield a correct measurement (see Vosmer & Engelen 2008). The crucial role of the interviewer makes interviewer training and instruction extremely important. See the subtheme ‘Interviewer training’ in Organization of Fieldwork.

It is noted moreover that a uniform stimulus (question text) is not always the best solution: see for example Houtkoop-Steenstra (2000), who advocates, within the tight constraints of standardization, allowing an interview to proceed in a way that bears some resemblance to a normal conversation. This is actually Statistics Netherlands practice. For example, it would be awkward if a respondent has already mentioned having a son and a daughter, to then proceed regardless of that remark to ask another two questions about the number and gender of any children. This approach would be contrary to conversational convention and the respondent might conclude that the interviewer was bored and not bothering to listen. It would be acceptable in the above example for the interviewer to enter answers to these two questions without asking. This approach would be unacceptable for more complex concepts: even if a respondent had already mentioned being unemployed, the interviewer would still have to ask all the follow-up questions in order to confirm that the unemployment corresponded to both Dutch and International Labour Organization definitions. See also Section 3.2.3. Routing and textual imputations allow questions to be made somewhat more ‘personal’. For example, depending on information revealed earlier in the interview, a question could be phrased in terms of ‘your spouse’, or alternatively ‘your partner’. This must happen consistently, which is why textual imputations are preferable to having individual interviewers improvise in this way. Automated textual imputations also simplify interviewers’ work.

For the sake of clarity, it is noted that a uniform stimulus does not mean that *all* respondents always have to receive exactly the same stimulus, only all respondents for whom a given target variable is to be measured. Different stimuli are used for measuring different aspects: you rap a wall with you knuckles to locate a hollow space, but you spin an egg to test if it is boiled. Routing ensures that respondents receive the correct uniform stimulus.

3.2.2 The response process

It is usually harder than the researcher thinks for a respondent to complete a questionnaire correctly. The response process model, which has its origins in cognitive psychology (e.g. Tourangeau, Rips & Rasinski, 2000) has four steps that a respondent may take in answering a question: comprehending the question, retrieving the appropriate information (from memory or documents), forming a judgment about the information, and responding. Few respondents will run through all the steps consciously. When asked ‘How old are you?’ most people will have an answer ready. For more complicated questions the respondent may need to reiterate some steps in the response process: for example, if the answer shows the respondent failed to comprehend the question properly and the interviewer provides some additional explanation of the question. Box 3.2 shows the steps in the process and gives an example.

Box 3.2 Example of a response process for population surveys

Survey question

The following questions are about contacts with a general practitioner (GP). They cover surgery consultations, GP home visits and consultations by phone. When did you last contact a GP on your own behalf? Was that:

- less than half a year ago;
- between half a year and one year ago;
- more than one year ago?

Steps in response process

Comprehension

The respondent is unsure whether phoning for a repeat prescription counts as a GP contact, since all it involves is leaving a message on the answering machine. On the other hand, the GP sends a bill for writing the prescription, and this leads the respondent to conclude that it constitutes a contact.

Retrieving appropriate information

The respondent tries to recall the date of the last contact with the GP from memory, but has forgotten. However, he knows it must have been just before the summer break, because the reason for phoning for a prescription was to have enough medication for the holiday. The respondent remembers that the holiday was some time in early July.

Forming a judgment on information

The respondent estimates from the approximate holiday date that the last contact with the GP was near the end of June. The respondent then works out that 'near the end of June' was less than half a year ago.

Responding

The respondent answers: 'less than half a year ago'.

Alongside cognitive aspects, organizational and administrative aspects also have a relatively important role in business surveys compared with population or household surveys. A somewhat modified model for the response process is accordingly used, derived from Willimack and Nichols (2001). See Box 3.3 for a summary of the response process for population and establishment surveys. There are three additional steps for business surveys prior to the cognitive process for individuals described above: information recording, respondent selection and prioritization by the organization. These three steps form the context for the cognitive process of completing the survey. Another additional step, information release, comes at the end of the response process. This reflects the fact that respondents in some companies first need permission from above before they can submit a survey form.

Box 3.3 Response process for population and establishment surveys

Response process	
Population surveys	**Establishment surveys**
	▪ Recording in administrative system/memory
	▪ Respondent selection
	▪ Prioritization
▪ Comprehension of question	▪ Comprehension of question
▪ Retrieving appropriate information	▪ Retrieving appropriate information
▪ Forming a judgment on information	▪ Forming a judgment on information
▪ Responding	▪ Responding
	▪ Formal release of answer

3.2.3 A questionnaire as a conversation

A questionnaire is an instrument in which the researcher has a conversation of sorts with the respondent. This is important to realize, because it means that account must be taken of conversational conventions in everyday life. For instance, we usually respond to the assumed meaning of a comment or question rather than the literal content. It would be perfectly normal to respond to the question ‘Are you coming along to the cinema?’ by saying ‘I have to work late this evening’. In a literal sense this does not answer the question, but everyone will understand that you cannot go.

Grice (1975) defines four maxims of successful conversation (see Box 3.4):

Box 3.4 Grice’s conversational maxims

Maxim	**Explanation**
Quantity	Give just the right amount of information for the purpose of the exchange.
Quality	Be truthful, say nothing for which you lack adequate evidence.
Relation	Talk about matters that are relevant for the purpose and subject of the conversation.
Manner	Make the contribution to the interview concise and clear.

Respondents assume that the researcher will have observed these maxims when drawing up the questionnaire (see e.g. Maynard et al., 2002). Respondents therefore expect the questions presented to be phrased so as to be easily understood and relevant. Deviations from the rule are perceived as disruptive: having already explained that they are on the permanent staff, someone would find it irritating to be asked later whether they had a temporary contract.

Respondents too will usually conform to the above rules. The maxim of quantity then means they will tend neither to give information that appears superfluous nor repeat what they stated earlier in the interview. Suppose a general question about satisfaction were to follow one about satisfaction with a specific aspect (e.g. work, marriage). The answer to the general question would then be influenced by the earlier question: the question about general satisfaction will be interpreted as 'general minus the aspect that was specifically identified' (Sudman & Bradburn, 1983, 143-144). Mechanisms of this kind also come into play when reporting 'hard' financial information. For example, if a business survey first asks about the costs related to temporary employees and then about labour costs, most respondents will report labour costs excluding costs related to temporary employees.

3.2.4 Four types of language in questionnaires

Often unconsciously, respondents will use all the information on a questionnaire, whether printed or on the web, when performing their tasks. The language used in questionnaires can be divided into four types: verbal, numeric, symbolic and graphical (Redline & Dillman, 2002; see Box 3.5).

A variety of experiments has shown that small changes in one of these language types can have a major impact on the way respondents answer questions (e.g. Dillman & Redline 2004). The different kinds of language must therefore reinforce each other, and must jointly support questionnaire answering. Consequently, the final questionnaire text and format must always be mutually compatible.

Box 3.5 Four types of language in questionnaires

Type of language	**Examples**
Verbal	Textual items such as headings, questions and explanations.
Numeric	E.g. numbering of blocks, questions, response categories and scale points.
Symbolic	Summation signs, arrows to the next question, a clickable icon for help text.
Graphical	Font size, position on the page, question spacing, use of colour.

It is important for a questionnaire to be internally consistent throughout in all types of language used. For instance, all explanations must be presented in the same style, and for any given concept (e.g. ‘employees on the payroll’) must always use the same words.

3.2.5 Important sources of measurement error

There has been much research into the measurement errors that occur when collecting data in standardized surveys. The most important sources of measurement error are described briefly below. Additional information is available in Tourangeau, Rips and Rasinski (2000) and Biemer & Lyberg (2003).

- ‘Satisficing’: running through the question-answer process in a suboptimal way

Respondents may sometimes be disinclined to go through much trouble to correctly understand and answer a survey question. They then go through the response process perfunctorily, or skip some steps altogether. For example, quickly guessing an answer, or opting for ‘don’t know’, to a question that actually asks for a calculation. The literature refers to this phenomenon as ‘satisficing’ (Krosnick, 1991).

- Socially desirable answers

Socially desirable answering occurs when respondents adapt their true answers to a question to what they think is socially desirable. This tendency may lead to either underreporting of behaviour (e.g. alcohol use) or over reporting (e.g. segregation of waste).

The literature shows that distortion caused by socially desirable answering plays a part in many subjects, including aspects that may appear at first glance to be relatively insensitive, such as perceived health problems or satisfaction with a neighbourhood. Researchers would appear to have some trouble determining how sensitive items are to socially desirable answering. Questions about the current situation appear to be more susceptible to distortion than those about the past (Fowler, Roman & Di, 1998).

- Acquiescence

Respondents would apparently prefer to agree with a statement, regardless of its content and despite disagreeing. The literature refers to this tendency as acquiescence. Attention must be given to balanced question phrasing to avoid this problem. An example would be: Do you agree or disagree with… .

- Memory effects

Memory effects may influence answers to questions that involve respondents’ long-term memory. Long-term memory does not store all information equally well, and information that is stored is not always readily recalled.

Clearly the most important memory effect is forgetting facts, which leads to underreporting. Another familiar memory effect is attributable to forward or backward *telescoping*, in which respondents incorrectly place certain events inside or outside the reference period to which the question relates.

Tourangeau, Rips & Rasinski (2000) gathered research findings about memory effects and the implications for questionnaire design.

- Events further in the past are harder to remember. Therefore restrict the reference period where possible in order to lower the frequency of observable events. For hospital admissions a reference period of one year would appear possible, while for GP contacts fourteen days is advisable.
- Events around important milestones in a person's life are easier to date. Memory can be stimulated by using these personal milestones, helping to order the events correctly in time. Accurate dating in this way first involves the interviewer and respondent jointly compiling a kind of calendar of important dates. The events measured by the survey are then placed on the calendar. The above approach is obviously possible only for interviewer-administered questionnaires. Web questionnaires would appear to demand additional research.
- Important and emotional events are easier to recall than less significant ones. For this reason you should use shorter reference periods for events of little interest to the respondent, at least if the frequency of the investigated events allows this. See also the first point.
- The use of various memory aids simplifies recall. Therefore use multiple types of memory aid (in the following order of preference): ask about the 'what' (type of event), 'who', or 'where' (the people involved or the location) and finally the 'when' (the period of time). Divide the recall task into components. For example, for frequent behaviour such as 'drinking', do not ask for totals for a whole day, but for consumption in different parts of the day.
- People recall more information if they spend more time on the task. Give respondents more time by taking longer to introduce the question, and slowing the pace of the interview (by speaking more slowly and allowing more time for reflection). The faster pace of telephone interviews makes this mode less suitable for difficult retrospective questioning.

▪ Context effects

Identical survey questions may be answered differently in different contexts. This effect could be attributable to the question sequence in a questionnaire, the format, the style of answer option (e.g. the number of character positions for an open question may suggest that either a long or a short answer is expected), or the title of a questionnaire, or the text of a covering letter. It is therefore important when comparing the results of different surveys to look closely at the context in which a question was set. Furthermore, when investigating trends, as few changes as possible must be made to the questionnaire: a change in question sequence is enough to evoke a different response, thereby reducing comparability. An obvious important context is the mode in which a questionnaire is offered. It matters whether a questionnaire is answered by the respondent on a computer, or administered by an interviewer at home. Mode effects occur mainly with cognitively demanding questions (which therefore demand a high degree of motivation) and those that are susceptible to socially desirable answers.

- Interviewer effects

The interviewer may have an unintentional but significant influence on how respondents interpret and answer questions. This may happen if an interviewer fails to administer the questionnaire correctly, for example by misreading the question and answer texts, or steering the respondent towards a particular answer. Sometimes an interviewer's personal traits will be enough to influence a respondent's answers. For instance, respondents may more readily report unfavourably about safety in the neighbourhood to middle-aged interviewers who themselves are anxious. Unless a question is completely unambiguous, the probability of interviewer effects will increase, since it will be up to the interviewer to find a way of clarifying the matter. The same happens with questionnaires that fail to take all options and situations into account, obliging the interviewer to improvise in order to proceed. Moreover, questions that often require additional explanation that is not given in the questionnaire itself in the form of an interviewer instruction will raise the probability of interviewers choosing interpretations of their own.

3.2.6 Special points for attention in business surveys

- Contacts change

Many Statistics Netherlands business surveys are sent with some regularity (e.g. annually or quarterly) to a fixed contact person in an organization. However, these contacts can change or go on holiday. The contact's successor or deputy will not necessarily have the same knowledge as their predecessor. The deputy may interpret concepts differently. All agreements with contact persons must therefore be recorded. All the information needed to complete a questionnaire correctly should normally be enclosed with the questionnaire, or at any rate be available in case the contact person changes.

- The questionnaire goes to multiple respondents in an organization

It is common in business surveys for a single respondent to be unable to answer all the questions. Sometimes parts of the questionnaire will be copied and forwarded to other departments. For example, questions about employees will be answered by HR, questions about operations by the accounts department, and so on. Any explanations given on a separate sheet of paper will not always be attached, which is why explanations of questions should always be positioned as near as possible to the question concerned.

- People reuse previous survey answers

Many companies archive copies of all completed Statistics Netherlands questionnaires. People will often consult the previous version when completing a questionnaire. Some companies provide links in their administrative system to support rapid generation of data for regularly recurring questionnaires. It would create problems for these companies if a questionnaire were suddenly to change. Frequent changes must be avoided, and changes that do occur must be marked explicitly.

3.3 Designing survey questions

This section discusses how to formulate questions and we suggest what you should consider in order to produce good questions. The guiding principle in drawing up questionnaires is that the interview is indeed a conversation, albeit within the constraints of standardization. See also Section 3.2.3. If a conversation of this kind is to flow smoothly, the questionnaire must be respondent-friendly. One important aspect of respondent-friendliness is only asking questions respondents can answer, and another is only asking about aspects that are relevant for, and perceived as important by them. Needless to say, a questionnaire should be as short as possible in order to collect the information needed efficiently. The above aspects must be taken into account when phrasing the questions. In this section we build on the more theoretical considerations from Sections 3.2.3, 3.2.4 and 3.2.5. See also Vosmer & Engelen (2008).

3.3.1 Only ask questions respondents can answer

Only ask about things respondents can reasonably be expected to be able to answer. Data that are hard to find in a respondent's head or files cannot be measured simply with a survey. Information of this kind could possibly be recorded in a logbook or diary kept specifically for the survey. Be prudent if respondents have to answer:

- proxy questions, which the respondent answers on behalf of someone else, who is the survey target person;
- questions about the distant past (e.g. all the vaccinations someone has had);
- questions involving much detail (e.g. a specification of the invoice value of all dispatched goods as part of the Road Transport survey);
- questions demanding specific expertise (e.g. about the competence of a respondent's GP);
- when probing motives and causes you should bear in mind that respondents sometimes are unable to describe exactly why they perform or omit certain actions. It is better to ask whether or not certain circumstances or motives exist;
- hypothetical questions (e.g. 'what would you do if ...') usually yield little reliable information.

3.3.2 Define the unit and reference period

Ensure that it is clear for the questionnaire as a whole and each question which unit and period the respondent is expected to report on. When answering questions it is important for people to have the subject clearly in mind. Is the survey form about the entire group or just one business unit? Is the questionnaire asking about the individual or the entire household? Is the question about the last employer to be mentioned, or the most important one? When Statistics Netherlands says 'usually', do they mean 'right now' or 'last year'? Every question for which it may be expected that the answer could vary with time must state a reference period (Fowler & Cosenza, 2008; see Box 3.6).

Box 3.6 Example of a reference period

> Not:
>
> 'How often do you feel depressed?'
>
> Instead:
>
> 'How often have you felt depressed in the past four weeks? Was that all the time, most of the time, sometimes, or never?'

The reference period of a question must be chosen with care, since it can both influence how the question is interpreted and hamper answering. If for example you were to ask how often someone had been angry in the past week, milder forms of anger would be reported than if the question had been about the past year. This is because less severe forms of irritation or anger will be more quickly forgotten than furious outbursts. See also Section 3.2.5. This explains why answers to questions about the same behaviour in different reference periods sometimes do not add up. Fowler & Cosenza give the following tips for defining reference periods.

- Define the reference period in relation to the importance of the event referred to in the question. The less important the event for the respondent, the shorter the reference period should be.
- Divide recall tasks into multiple steps. For example, rather than asking directly about how many contacts someone has had with doctors in the past year, ask about several specific kinds of doctor.

3.3.3 Ask one question at a time

If you ask more than one question at a time, it will be unclear which question the answer refers to. A sign of a double question is the presence of words such as 'because', 'or' and 'and' (see Box 3.7).

Box 3.7 Example of a double question

> Not:
>
> 'Are you economical with energy because of the costs?'
>
> Instead:
>
> 'Are you economical with energy?'
>
> If so:
>
> 'Is that because of the costs?'

3.3.4 Phrase questions in a balanced way

Respondents often tend to be compliant in their answers (the 'acquiescence' referred to in the literature). This form of distortion can be reduced by making questions as balanced as possible. Therefore, rather than asking 'Do you agree with raising the pension age from 65 to 67?' ask 'Do you agree or disagree with …'. See also Section 3.2.5 for a discussion of this measurement error.

3.3.5 Avoid making false assumptions in phrasing the questions

A familiar source of problems in survey questions is basing the design on false assumptions about the respondent's knowledge or conduct.

To ask without further explanation for an opinion about some fact (e.g. 'the new medical expenses scheme') is to assume that the respondent knows what that is. Several experiments have shown that many respondents will happily answer questions about subjects they know nothing about. They may do so because they are unwilling to admit never having heard of something or because they misinterpret the question and are then able to answer in a way that makes sense to them. See Box 3.8 for an example of this effect.

Krosnick et al. (2002) conclude from various studies that solicited respondents' opinions about fictional matters (e.g. a nonexistent regulation), that approximately 30% of respondents will answer this kind of question despite actually being in no position to do so. If it is likely that many respondents will not know exactly what a question is referring to, ways of avoiding problems would include providing an extensive introduction to the subject, inserting a filter question ('Are you familiar with …'), or providing an explicit 'don't know / no opinion' option.

Box 3.8 Example of interpreting an unclear concept

A Statistics Netherlands experiment (Vis-Visschers, Arends-Tóth, Giesen and Meertens, 2008) into the use of 'don't know' and explanations with Internet questionnaires deliberately used difficult words when phrasing the questions. One of the questions was: 'Do you or anyone in your family ever eat products that contain retinol?'

A button was provided for an explanation of the term retinol. Relatively few respondents used it.

One of the respondents who claimed never to have eaten retinol was an interesting case. When asked to explain his answer, it appeared that he had incorrectly assumed that retinol must be something unhealthy, and that his wife would make sure he did not eat any.

Besides misjudging the level of knowledge, survey questions are sometimes based on false assumptions about respondent behaviour. For example, a question that asks parents of young children how often their child wears a helmet while being carried on the bike prematurely assumes that the child sometimes travels by bike. If this is not the case, and there is no response category for ‘not applicable’, respondents have no way of answering.

3.3.6 Choose clear words; provide any necessary definitions

Valid and reliable measurements depend on the use of unambiguous concepts. Even seemingly simple concepts can leave scope for disparate interpretation. For instance, the term ‘last week’ could be interpreted as ‘from Monday to Sunday, inclusive, of the previous week’, or ‘last Monday to last Friday’, or, if the question was asked on a Thursday: ‘last Thursday to last Wednesday’. All abstract and ambiguous concepts must therefore be defined. Many respondents in cases of this kind will assume that their own interpretation is the only right one and therefore will not think to ask what was actually meant.

Definitions and explanations are preferably given prior to the question, or incorporated into the question text. This clarification will immediately give respondents the information needed when interpreting the question, and will avoid having them interrupt or stop the interviewer reading the question to the end. Never place an intermediate question mark in a question text, because that triggers interruptions.

Sometimes feeding respondents with a prescribed definition will be deliberately circumvented. The reason could be that the Statistics Netherlands definition deviates too much from that of respondents, or that individual respondents differ too much in their views about the concept. An example would be the definition of part-time work. Respondents have widely varying ideas on this subject, and the Statistics Netherlands definition (less than thirty hours) is definitely not endorsed by everyone. It would be better in this case to ask about the number of hours a person is employed and then deduce whether this constitutes part-time or full-time.

Otherwise, it could be an explicit research goal to explore respondents’ understanding of a given concept, in which case it would be appropriate to give no definition.

3.3.7 Use simple language

Statistics Netherlands surveys are in general intended for a broad group of people, making it important to use simple language. The best advice is to use language at proficiency level B1 (see Appendix 1 for definitions of the different language proficiency levels). Almost everyone is able to understand language level B1, even if they are not highly educated and never need to read in their work. The most important tips for writing simple language are given below. There is information about writing simple Dutch in the Schrijfwijzer begrijpelijke formulieren (Bureau Taal, 2006) or on the website of the City of The Hague. This website has many writing tips and even has a dictionary of difficult words and alternatives (‘Het Helder Haags Woordenboek’).

- Keep questions and sentences brief.

Brief questions are easier to understand and remember than long ones. However, short is not always better, and sometimes clear understanding demands an explanation, while a longer introduction to a question can actually give respondents an opportunity to focus on the relevant subject or period. Conversely, excessively long introductions or explanations can overshoot the mark if they are too much for respondents to remember. Always aim to keep the sentences in a question text as short as possible. As a guide, twenty-five words a question is good (Vosmer & Engelen 2008).

- Use commonly used and well-defined words.

Jargon, figurative use of language and abstract terms are hard for many respondents to understand. The target populations for Statistics Netherlands business and population surveys alike are often extremely heterogeneous. The use of language must also be comprehensible for the subpopulation with low language knowledge.

- Texts containing double negatives are often hard to understand.

'I disagree with you not doing that' is already hard enough. Double negatives in survey questions often arise in statements asking respondents to indicate their degree of agreement or disagreement. Negatively formulated statements then create a double negative. Because what does 'disagree' mean as an answer to the statement 'I am dissatisfied with my work'? Respondents who are satisfied with their work are supposed to answer *disagree* here. Mistakes are easily made.

- Phrase question texts and instructions positively.

Positively phrased texts ('use a pen') are easier to understand than negatively phrased texts ('do not use a pencil'). Beware of using words such as 'not' or 'excluding', which respondents often overlook.

3.3.8 State the required answer format clearly

Be as clear as possible in the question itself as to what kind of answer is required. It will then be easier for the respondent to answer. This point applies in particular to interviewer-administered questionnaires, where the respondent usually does not see the answer options (see Box 3.9).

Box 3.9 Make the answer format clear in the question

> Not:
>
> 'When did you move here?'
>
> Instead:
>
> 'In what year did you move here?'

3.3.9 *Questionnaire design for international use*

Researchers will often want to compare the output of research from different countries. This places special requirements on questionnaire design, since there will be a need to measure as closely as possible the same things in each country. Various procedures have been devised for harmonizing concepts and questions in an international context. See for an extensive treatment of this subject Chapter 4 in Brancato et al. (2006).

3.4 Design of answer options

A wide variety of factors has to be considered when designing answer options that are presented to respondents after asking a survey question. The most important are the following.

- Level of measurement required

  The kinds of analyses that can be performed on the data depend on the level of measurement of the variables. The required level of measurement determines such things as whether age or income is measured in classes rather than in years and Euros. Response categories must therefore be chosen in consultation with the envisaged data users.

- Costs for processing and data collection

  Closed questions are easier to process than open ones. Closed questions have fixed answer options, whereas open questions leave respondents free to formulate their own answer. Coding the answers to open questions is a time-consuming process that cannot be performed completely by software. It is likewise time-consuming to have interviewers note down the answers to open questions.

- Ease of answering

  The choice of the kind of answer options influences how difficult a question will be to answer. It is easier to give a rough indication of which of five income classes you fall into than to state your annual income accurate to the euro.

- Quality of the answer

  Respondents use the answer options in interpreting the purpose and meaning of a question and will glean from them the researcher's evident expectation of the distribution in the population. Respondents may adapt their answers to the answer scale because they want to give a socially desirable answer. On the other hand, they can use the scale to arrive at a quick answer: the midpoint on a scale is often interpreted as an average answer. If someone thinks they perform a certain activity somewhat more often than average, it is easier quickly to pick an answer option above the middle than first to determine the precise answer and then match it with the answer scale. See Box 3.10 for an example of the effect of answer scales.

Box 3.10 Effect of different answer scales

For how many hours a day do you watch television?

Option 1	≤½	½ - 1	1-1½	1½-2	2-2½	> 2½
Option 2	≤ 2½	2½-3	3-3½	3½-4	4-4½	> 4½
Option 3	Open question					

% of respondents that watch more than 2.5 hours of TV a day:

Option 1	22%
Option 2	54%
Option 3	52%

Source: Toepoel, Vis, Das and Van Soest, 2009

Table 3.1 lists common answer options and gives some tips about how to use them. The list is based in part on Jansen et al. (1989). Details of all the factors to be considered in the development of response categories are outside the scope of this Methods Series. A good overview is given in Chapter 3 of Brancato et al. (2006) and in Billiet (2006). The latter is in the Dutch language and has many examples.

Table 3.1. Kinds of answer options

Answer option	**When to use**	**Tips**
Open The respondent is required to enter data in their own words, codes or numbers	- If the number of possible answers is large or the answers to be expected are unknown; - provided it is clear to the respondent what kind of answer is expected.	Elicit an answer that is precise and to the point: - make the question as specific as possible; - start the question with a question word (e.g. how, who, which); - do not make answer boxes too large.
Half open Some fixed answer options are provided, plus an answer option 'other, i.e. .....', which opens the possibility of giving an answer that is not in the list.	- If the most obvious answer options are known but there is some doubt as to whether all the relevant options are covered.	Use this option for new questions. If the question has been in the field for some considerable time, some frequently occurring answers may be considered for inclusion as separate answer options. The less frequently occurring can then be noted in the closed category 'other'.
Single choice The respondent has to choose one of two alternatives, e.g. yes or no, larger or smaller.	- If the applicability of one specific category is important.	- Consider whether the question is always relevant, or that a 'not applicable' category should be added.
Multiple choice questions The respondent must either choose one answer from various options, or mark more than one answer.	- If what is important is which of a series of categories is applicable; - if the number of possible answers is limited; - N.B. if more than one item in a list may be applicable, the alternative of including a separate yes/no question for each should be considered. This will be more work for the interviewer and respondent,	- make sure the list is exhaustive; - make sure the answer options do not overlap; - make it clear whether one or more answer options may be marked; - if answer option X is a special case of Y, place X before Y; take care that the response categories in cases of this kind are formulated so as to be mutually exclusive; - if a selection has to be made from a long list, make sure the sequence will be helpful to the

Answer option	When to use	Tips
	but will usually give a higher (and better) reporting of the investigated aspects.	respondent (e.g. alphabetic, geographical); - in the absence of other considerations, place the most likely answer option at the top; - it is preferable to mark what is applicable (rather than crossing out what is not); - preferably do not ask for codes to be entered; - long lists are hard to process, so if possible break long lists using subheadings.
Answer scales	Often used when respondents are asked for their opinion.	- It is usual to have scales with between 4 and 7 points, where the number of points depends on analysis requirements; - the ease of completing a questionnaire decreases as the number of points in a scale increases; - have as many negative as positive points in the scale; - for scales with up to 7 points label all points with words; - put the neutral option at the midpoint of the scale; - use the same scales wherever possible within the same block, and preferably also throughout a questionnaire; - endless lists of scale questions are tedious to answer and lead to suboptimal completion behaviour; - preferably do not use answer scales for indirect assessments (see below).

A brief description of the use of answer scales for indirect assessments is given below. Fowler & Cosenza (2008) differentiate between the way answer scales are used in direct and indirect assessment tasks, and advise against the use of indirect assessment tasks. A direct task is one in which respondents are asked to place their opinions about something (e.g. their own health) on a continuum that is concerned with this aspect (e.g. ranging from bad to excellent). An indirect assessment task is one in which respondents are presented with a statement (e.g. ‘my health is excellent’) and then asked to indicate on a continuum their degree of agreement. Indirect assessment tasks are far more complex and therefore more susceptible to measurement error.

3.5 Use of explanations and instructions

Explanations and instructions are important tools for clarifying to respondents the substance of a question and how to answer it. With population surveys in particular, it is doubtful whether a question can be considered good if it needs much special explanation and clarification. The situation is less clear cut with business surveys, in that accurate specifications are often asked for, and do not hamper answering. Unfortunately, explanations in self-administered questionnaires are often poorly read. This may be detrimental to comparability if some respondents but not others read the explanation. In general, the text that is read best is the question. Therefore, any essential information that all respondents must be aware of should preferably be included in the question text. The most important guidelines for using explanations and instructions are as follows.

- Factors to be considered in positioning explanations about the meaning of a question or certain concepts:
  - first choice: in the question text itself, before the actual question;
  - second choice: short and to-the-point in a separate explanation, near the question or answer space;
  - third choice: in an extensive explanation at the bottom of the page, or clickable from the question as an instruction for the interviewer;
  - fourth choice: as a separate document. Individual questions in the questionnaire must provide clear references to any explanation in a separate document.
- Place general background information about the survey in an accompanying letter or in separate tips for answering. Other possibilities include a folder or a web site.
- Instructions for completion that clarify how to answer the questions (e.g. round to the nearest thousand) must be integrated into the question, immediately above or alongside the answer box. An instruction that applies to most of the questions may be placed at the beginning.
- Keep explanations as short as possible.
- Do not use long sentences in questionnaires that are to be self-administered by respondents.
- With CAWI questionnaires, position any error messages where they occur (and preferably not somewhere at the end of the form) and in such a way that the original question remains visible to the respondent. Formulate error messages in a friendly and clear way so that respondents will understand what they have done ‘wrong’.

3.6 Guidelines for questionnaire structure

Satisfactorily administering or completing a questionnaire is a complex task. Dividing it into small, manageable subtasks makes the work considerably easier for respondents and interviewers. The subtasks must be presented in a logical sequence and – if necessary – subdivided into thematic blocks. Important points for interviewers include knowing exactly how far through the questionnaire they are, which was a simple matter with the once-common paper questionnaires. To simplify this for computer-assisted questionnaires, the lower frame of the screen shows the preceding questions with the answers given, and a few subsequent questions.

The following is mainly about questionnaires that are completed by respondents themselves. A satisfactory structure provides something to hold on to before starting (what is it about, how is the form structured, and who should provide the information) and while answering (e.g. how a given question relates to others). A way of doing this with computer-assisted questionnaires is to display an index, or multiple tab sheets.

It is important that the structure and format of a questionnaire is designed such that respondents can always easily find their way through and can see where they are in the process. When reopening a CAWI questionnaire this could be achieved by displaying the screen with the last question that was answered. Self-administered questionnaires, for business surveys in particular, are usually worked on in spare moments, with occasional interruptions. Tools of this kind are even more important in this situation.

3.6.1 Naming of questionnaires and thematic blocks

Give every questionnaire a clear name. The position, layout and size of the name make clear that the information applies to the entire questionnaire.

- Keep the name as short as possible.
- Choose names that mean something to respondents.
- The names of questionnaires that are issued periodically to the same respondents must also indicate the period to which the questionnaire relates.
- Choose a unique questionnaire title.
- For self-administered questionnaires preferably also number questions and blocks.
- A questionnaire name or the name of a specific block of questions tends to be read only when browsing through before starting, but skipped when actually answering the questions. It must therefore also be possible to answer the questions correctly without reading the question block names.

Box 3.11 Example of interpreting questionnaire names

> A draft questionnaire had the title *'Quarterly Company Census'*. This is a survey conducted by Statistics Netherlands four times a year, but with new respondents each time. Under these circumstances, it is unhelpful to respondents to use the qualification Quarterly, and it sometimes even prompts questions like 'does that mean I will have to complete it every quarter if I take part?'

3.6.2 Subject sequence

- The initial impression sets the tone. Therefore start with an easy and interesting topic, and preferably not with an interrogation about miscellaneous background characteristics (although this is sometimes unavoidable if the information is needed for questionnaire routing).
- Group questions that belong together. Try to avoid making respondents jump back and forth between different topics, reference periods and answer options.

- Place sensitive subjects at the end of the questionnaire, to be answered after the respondent has established a bond with the interviewer and the survey. Should the sensitive subjects lead to terminating the interview, then at least many of the questions will have been answered.
- Do not place any difficult questions that demand much effort and motivation too far down the questionnaire, when fatigue may have set in. On the other hand, it has been shown that extremely difficult questions lead to termination. Clear dropout peaks coincide with questions of this kind in web questionnaires.
- Questionnaire subjects may influence each other. Take this possibility into account in the question sequence. For example, do not start by asking if someone is unemployed and move on immediately to asking for a general opinion about the economy.

3.7 Questionnaire format

Good questionnaire format is important for both ease of answering or administering the questionnaire and the quality of the answers. Much has been written about questionnaire format. It is possible in this section only to provide a brief impression, and what follows is largely based on Chapter 3.3 of the *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System* (Brancato et al., 2006). Other useful sources with additional information about questionnaire format are Dillman, Smyth & Christian (2009) and Couper (2008).

People respond to format on three levels: emotional, functional and reflective. Responses on the *emotional level* are the initial, impulsive feelings evoked by the questionnaire's appearance. A thick questionnaire crammed with tiny letters will create a different impression from a thin counterpart laid out neatly and legibly. An attractive questionnaire format is not to be dismissed as merely cosmetic, and should not be skimped. Questionnaires with an appearance that immediately evokes a bad feeling will be completed less often and less well. Some features of attractive format are symmetrical, rounded shapes, recurring patterns and plenty of white space.

On a *functional level* the format must support respondents and interviewers in their cognitive task. For instance, it is useful for subtasks also to be visually distinct, for it to be clear which answer boxes belong to which question (clear leaders), and for any long lists of answer options to be divided into clear chunks. It is also important to align as closely as possible with existing practice. Some familiar recommendations for good functional design are:

- people in our culture read from top left to bottom right: a questionnaire must therefore start at the top left;
- aspects that belong together (e.g. questions within a thematic block, answer options for a question) must also be clearly visually clustered;
- information that is irrelevant for the interviewer or respondent (e.g. intended only for internal processing) must be less visually conspicuous. This

information should therefore be in a smaller font and preferably at the bottom of the page;

- there must be a clear visual indication of where the answers are to be entered, e.g. by using a light-coloured background with white answer boxes;
- the format must always be functional and consistent: use the same font sizes, letter types, colour and graphical symbols for the same components (e.g. block titles, answer options and explanations);
- colours should serve only to supplement the layout, and never be used as a means of meaningless variation. Bear in mind that some people cannot see all colours well, besides which you can never know how the questionnaire will be shown to a respondent or interviewer (e.g. a paper questionnaire can be photocopied, or web questionnaires are shown differently under different browsers). Take particular care with reds and greens, which many people distinguish poorly.

On the *reflective level* a respondent will consciously deliberate on the response task, possibly deciding whether or not to answer the survey honestly, or to take part at all. It is important on this level for the questionnaire format to emanate a serious attitude on the part of the organization conducting the survey. Relevant items are the use of the Statistics Netherlands logo and a restrained yet polished format.

4. Methods for developing, testing and evaluating questionnaires

4.1 Introduction

It is almost impossible to develop a good questionnaire in a single session from behind a desk. Several properly devised pre-tests can help avoid many problems in the collection and analysis of the data, although pre-tests are unable to highlight and resolve all the possible issues with questionnaires. It is therefore useful to evaluate questionnaires that have been in the field, facilitating ad hoc intervention, and the extraction of supplementary information for analysing the data and possibly adjusting the questionnaire in the future.

Chapter 2 outlined the process of questionnaire development and briefly presented the possible methods for testing and evaluating questionnaires. This chapter extensively describes several of the methods mentioned above. The methods described are used regularly in Statistics Netherlands and other statistics institutes and may be considered best practices or recommended practices (Brancato et al. 2006, Chapter 6). This chapter covers informal tests, expert review, focus groups, cognitive interviews, vignettes, observations, technical tests, experiments, and the analysis of process and survey data. A chart is first given of which tests are applicable in which phases of the process (see Section 4.2), before presenting the methods in detail: the purpose, any phasing, and advantages and disadvantages. The most important sources used in writing this chapter were Brancato et al. (2006), Jansen et al. (1989), Snijkers (2002) and De Leeuw, Hox & Dillman (2008).

The most appropriate method depends on the purpose of the test, and in practice also on the time and resources available. A Eurostat minimum requirement for questionnaires is that they must be tested at least once with potential respondents (Brancato et al., 2006, page XI). It is efficient to employ the methods in order of increasing expense: for example, it would be a shame to give a questionnaire with blatant errors to test persons.

4.2 Which method for which phase?

Table 4.1 shows a chart of the methods that may be applied in one or more phases of the questionnaire development process. The phase or phases in which each method can be used is noted.

Table 4.1. List of methods for each questionnaire development phase

Phases / Methods	Conceptualization & research design	Questionnaire design	Pre-test	Revision	Data collection	Process-monitoring & evaluation
Document review	X	X	X			
Informal test	X	X	X	X	X	X
Expert review	X	X		X		X
Focus group	X	X	X			X
Cognitive interviews	X	X	X			X
Vignettes	X	X	X			
Observations			X		X	X
Pilot			X			
Technical tests			X			
Experiment			X		X	
Process data analysis					X	X
Survey data analysis					X	X

4.3 Informal tests

Brief description: Informal tests are those performed by the actual questionnaire author to evaluate a questionnaire.

Purpose: 'Quick and dirty' tests can be performed to gain a quick impression of the quality of the phrasing of questions, or to identify possible pointers for follow-up tests. These methods are extremely effective for rapidly tracing difficulties in the question-answer process, unclear aspects of the concepts and terms used, and of the length of question texts and questionnaires. They can also be a fairly simple way of revealing sloppy phrasing, poorly flowing sentences, difficult use of language, typographical errors and layout technicalities. This method can also be applied when time for large-scale testing, or for test preparation, is scarce.

Phase: This method is actually applicable in every phase of questionnaire development, but is used mainly at an early stage of conceptualization and questionnaire design.

Extensive description: One of the first things a questionnaire author should do is personally read the question and answer options out loud (Campanelli, 2008). A question that looks good on paper can still sound strange. Furthermore, it is worthwhile personally making a serious attempt to answer the question, based on the author's own situation or a few imaginary situations (e.g.: 'how would a self-employed person answer this question?', 'what would someone who has dropped out of an educational programme enter here?'). This method may appear trivial.

However, if all questionnaire developers were to try seriously to answer their own questionnaires just once, a mass of irritating questionnaire errors would quickly be revealed. It can also be worthwhile to engage family, friends and colleagues as kinds of test respondent: different people have different perspectives and also find different errors (see Box 4.1). For instance, questionnaire developers can hardly be expected to observe their own unconscious use of jargon when phrasing a question. Having questionnaire developers critically run through their own questionnaire using a checklist can also be viewed as an informal test method. There are numerous checklists for good questionnaires (see Appendix 2 for an example).

Box 4.1 Example of a 'quick and dirty' test

The question about GP contact was presented to a friend in order to find out early in the test process whether it might give problems.

Question: How often in the past 4 weeks, in other words since {date: 4 weeks ago}, have you contacted a practice assistant and/or specialist nurse in your GP's practice on your own behalf?

Response: 'No idea. What's a practice assistant? Is that the doctor's assistant? If so I would say yes, because she wrote my repeat prescription.'

Comment: a practice assistant is very different from a doctor's assistant.

Advantages and disadvantages: Informal testing can be organized relatively quickly. The findings are indicative of the problems a questionnaire could throw up, even if little time is available for testing. Needless to say, the most important disadvantage is that no 'real' respondents are consulted, and only the perspectives of questionnaire authors and people they know are involved in the test.

4.4 Expert review

Brief description: An expert review is a form of testing in which experts are consulted to give feedback on a draft questionnaire.

Purpose: The purpose of an expert review is to screen a draft questionnaire for potential pitfalls for interviewers and respondents. It is also possible to check individual questions, the operationalization of concepts (how concepts are translated into specific questions) and the design.

Phase: An expert review would usually be held in the conceptualization and questionnaire design phases. Findings often yield points for attention that can be used as input for the test phase in the questionnaire development process.

Extensive description: A variety of experts can be consulted, and which would be best obviously depends on the questions the questionnaire developers are interested in. If it is unclear whether the concepts have been correctly made operational, it could be worth consulting potential users of the data. It is useful to seek the opinion of questionnaire experts if there are questions about how comprehensible the questions and the questionnaire structure are. Some relevant kinds of experts are:

- content matter experts / statisticians;
- questionnaire methodologists;
- designers;
- envisaged processors of the data (both technical and subject matter aspects);
- help desk staff;
- interviewers and field staff.

Expert reviews address issues such as the terms and concepts used in the questions, question structure, response categories, question sequence, questionnaire navigation structure, and interview and coding instructions.

There are numerous ways of setting up an expert review. A way that works well is to have the experts assess the questionnaire systematically in advance, and arrange a later group discussion, or personal interviews, about the findings (see Section 4.5). An example of an extremely detailed assessment scheme is given in Appendix 3. A simpler way is what is known as the plus-minus method, in which the expert is asked to place a plus or minus sign against any (possibly previously identified) question that has some conspicuously positive or negative aspect. An opportunity is given to explain these pluses and minuses in the discussion.

Box 4.2 Example of an expert review

The evaluation of the Production Statistics (PS) questionnaires in 2004 involved many activities intended jointly to provide a clear picture of any points for improvement. One of the activities was to ask the opinion of Statistics Netherlands employees with experience of working with PS questionnaires (e.g. field consultants, help desk staff, data editors and analysts). They are experts in their field.

Statistics Netherlands employees formed focus groups that listed the problems with the PS questionnaires. All discussions had the following 3 components:

1. a general part, with problems that respondents have with all Statistics Netherlands questionnaires;
2. general problems with the PS questionnaires;
3. a specific part, with problems with the specific sections of the PS questionnaires.

The focus groups yielded many comments, leading to the conclusion that the questionnaires needed drastic revision. The comments gave pointers to the kinds of changes the questionnaires needed.

Advantages and disadvantages. Compared with surveys involving respondents, an expert review is relatively quick to implement, in particular if the experts are internal employees. Experts are usually well able to suggest workable solutions for problems that are identified. Compromises between multiple interests must often be made when creating questionnaires. It will help generate support for these compromises for all involved parties to be consulted as closely as possible in questionnaire creation. However, an expert review is only as good as the experts consulted. And even the best experts have a less than perfect understanding of respondents' problems and abilities. For an example of an expert review, see Box 4.2.

4.5 Focus groups

Brief description: A focus group is a structured group discussion usually involving between 5 and 7 people, led by a discussion leader (known as a moderator).

Purpose: Focus groups are used to obtain qualitative information about a certain target population's attitude to predefined topics, e.g. the subject of the questionnaire, the envisaged mode of data collection, or the concepts and definitions used in the questions.

Phase: Focus groups can be applied both during questionnaire design and when evaluating an implemented questionnaire. A focus group is a convenient way of evaluating a questionnaire with, say, a group of interviewers or help desk staff.

Extensive description: Focus groups are particularly useful for gaining an understanding of perceptions, opinions and attitudes. There are two advantages of a group discussion compared with a one-to-one interview. The first is that it is more efficient to talk with six people together than one at a time. The second is that participants are encouraged by the group dynamics to be more outspoken and provide more detail when expressing their opinions and experience.

Among the subjects that could be raised in a focus group with test respondents are the sensitivity of questions, the interpretation of questions, the use of words and how easy or difficult it is to answer the questions. For an example, see Box 4.3.

A focus group demands thorough preparation if it is to be successful. A 'topic list' or 'topic guide' (a list of points for attention for subjects to be raised in the discussion) is essential for providing direction to the discussion, and the moderator must be constantly alert in guiding the discussion so as to serve the listed subjects. A topic list can be developed based on the output of an earlier review of existing documents, or an expert review. It is recommended to hold several group discussions involving between 5 and 7 participants in order to guarantee a satisfactory spread of experience and insights. Participants in the same group must not differ too much, or the discussion structure could suffer. The first task of the moderator is to make clear to participants what is expected of them (at any rate to divulge their insights and experience about the requested subjects and to treat the other group members with respect), the second is to monitor the topic list, and the important final task is to keep dominant speakers in check and to give quiet people a chance to speak.

A focus group will take approximately 1 to 1 ½ hours. It is very advisable to have an observer or minute-taker attend the focus group in support of the moderator. A video recording will also help in the analysis of all that was said (it can be difficult to determine who said what with an audio recording). Video stills can also be used as illustrations in focus group reports. There is additional information about focus groups in Krueger (1994).

A special style of group discussion is the kind of interviewer evaluation performed in Statistics Netherlands (see the description in Appendix 4). This involves a group of approximately 15 interviewers. The interviewers then divide into pairs and formulate statements about the questionnaire under evaluation. The statements may be about complaints and lack of clarity encountered in the execution of fieldwork, or, conversely, conspicuously favourable points. When complete, the pairs' lists of statements are put on display. An explanation of the statements may be given, but discussion is not permitted. The output of the discussion is a list of clear statements. At the end of this phase any missing subjects that were suggested in advance by Statistics Netherlands or the customer are identified. Any missing statements are added to the full list and explained as necessary. All interviewers are finally requested to give each statement a numerical score on a scoring form. The interviewers' scores facilitate a weighted expression of any obstacles. This method therefore offers little scope for discussion and exploration of perceptions, but is a quick and efficient way of listing perceptions and opinions.

Box 4.3 Example from the Family Monitor focus group topic list

This focus group with test respondents starts by viewing a video of a specimen interview.

Discussion about the video. Important: the discussion is not primarily about the couple's opinions, but the individual's!! Any time the woman says something about 'us' or 'we', the man should be asked what he thinks about it, and vice versa.

- What do you think about interviewing family members separately?
- What do you think about having the other person present or absent during the interview of the individual concerned?
- What would you prefer: the other person to be present or absent?
- Would this be feasible for you at home considering the layout of your house?
- And what do you think about confronting the couple with each other's answers?
- Would you personally be willing to take part in a survey where you and your partner were interviewed separately?
- And would you personally participate in a confrontation?
- Do you think it would be feasible to conduct a survey in this way with a large number of Dutch people?

Advantages and disadvantages: Group discussions are an efficient way of mapping out multiple opinions and perceptions at the same time. The group dynamics may encourage participants to formulate their ideas more clearly. On the other hand, social interaction can sometimes inhibit participants from expressing unusual ideas, or even from speaking at all. Focus groups with respondents to consider draft questionnaires will never be more than theoretical unless they are combined with serious attempts to complete or answer the questionnaire. Problems with questions often reveal themselves only under these circumstances. It must be observed finally that some people doubt the value of focus groups about hypothetical behaviour (‘what would you do if...’). While focus groups are useful in *understanding* what people take into consideration, the predictive value of discussions for future behaviour is usually meagre.

4.6 Cognitive interviews

Brief description: Cognitive interviews are open interviews with the purpose of gaining insight into the cognitive processes triggered by answering survey questions.

Purpose: The purpose of cognitive interviews is to gather qualitative information about how respondents understand and answer questions. the verbal utterances of test respondents are assumed to be direct expressions of the specific cognitive processes that occur in response to the questions in the questionnaire. The idea is that understanding respondents’ underlying cognitive processes may contribute to questionnaire design, the choice of response categories, question sequences, visual design, and so on.

Phase: Cognitive interviews may be held as soon as draft questions are available, and in the test phase, but may also be used to evaluate questionnaires already in the field.

Extensive description: Respondents in cognitive interviews are asked to describe their thinking as they answer survey questions. Observation and discussion of test respondents’ answering behaviour reveals problems with the questions and the answer options on the form. Below we first describe the techniques used in cognitive interviews and then discuss the practicalities involved.

- Techniques for gaining insight into cognitive processes

  We discuss below two frequently used techniques for gaining insight into respondents’ cognitive processes when answering survey questions: ‘thinking out loud’ and asking meta-questions. Vignettes can also be used to gain insight into cognitive processes. However, since this method is also suitable for other purposes, it is treated separately in Section 4.7. For an extensive treatment of cognitive interviewing see Willis (2005).

  <u>Thinking out loud</u>

  In the thinking out loud technique test respondents are asked to think out loud, expressing any thoughts that come into their heads while giving an answer to the question. Difficulties may be identified in comprehending the question, the response task, retrieving memories and facts from the past, responding to sensitive questions, and so on (see Box 4.4). An advantage of this method is that the researcher exercises no control, allowing unexpected problems also to emerge. A disadvantage of the method is that it demands appreciable verbal skill of the respondent, and the thinking out loud can actually disrupt cognitive processes. A respondent may then either be distracted from the task, or, conversely, fulfil it more conscientiously, either of which introduces artefacts into the testing. Some problems may arise in testing that would not occur in a normal interview, and others could remain concealed because test respondents try much harder than normal when completing the questionnaire.

Box 4.4 Example of thinking out loud in the preparations for the 2004 Family and Fertility Survey among Young People from Ethnic Minorities

> Response to the question: 'How often in general do you go to church or a religious gathering?':
>
> 'How can anyone answer this question properly? I'm a Muslim so I don't go to church. What a strange question. But I sometimes visit churches to satisfy my curiosity and to see why my Christian friends go there. I want to know. But how should I answer the question? Should I put down my visits to churches, or to my mosque?'

Meta-questions

Meta-questions are supplementary questions inserted between the survey questions (see Box 4.5). The purpose of meta-questions is to gain insight into respondents' thought processes. Some examples of meta-questions are 'What came to mind when you were answering this question?' or 'You answered yes. How did you arrive at this answer?'. Meta-questions can be devised either prior to the interviews or asked spontaneously ('You sighed when you came to that question. Why was that?'). Some special forms of meta-questions are as follows.

Probing or *persistent questioning*. In this technique a test respondent's answer to a survey question will prompt supplementary questions for clarification. These meta-questions are used if the information that a test respondent gives is incomplete, or where problems are suspected with a question. The interviewer may ask how a test respondent arrived at an answer, or why a particular answer option was chosen.

Box 4.5 Specimen survey question and associated meta-questions

Introduction: The following questions are about hospital admissions. They are concerned with hospitals of all kinds. Visits to the emergency department or outpatient clinic must not be counted.

Question: Have you in the past 12 months, in other words since {date 1 year ago}, been admitted to hospital for day treatment, i.e. an admission in which you were given a bed, but did not have to stay overnight?

Meta-question: This question was about admission to hospital. What hospital came to mind when you answered this question? What do you understand by 'hospital admission' and what by 'day treatment'?

Paraphrasing: Respondents in this technique are asked to restate the question in their own words. This meta-question can help confirm whether the test respondent has understood and interpreted the question correctly. Paraphrasing can also yield better words for definitions and concepts if multiple test respondents consistently use the same terms. This technique is particularly useful with complex questions and to identify jargon.

Some examples of the use of paraphrasing are: 'Can you restate this question in your own words?', 'Suppose you had to ask your next-door neighbour this question in your own words, how would you phrase it?'.

It is important when formulating meta-questions to have already some idea or hypothesis about possible problems with the survey questions. This information could have emerged from earlier focus groups or reviews. Ideally a test objective would be defined for each meta-question, so that the interviewer always has a clear idea of the purpose (see Box 4.6).

Meta-questions may be set at various times: immediately after a survey question, retrospectively after a block of questions, or after completing the entire questionnaire. An advantage of asking meta-questions immediately is that respondents will have the survey question and their response fresh in their minds. A disadvantage is that meta-questions disrupt the normal response process, thereby introducing artefacts (see also thinking out loud). From this perspective, it may be useful to ask some of the test respondents meta-questions immediately, and to question others retrospectively.

Meta-questions are another example of a cognitive technique that depends for its effectiveness on the skills and capacity of test respondents in verbalizing a train of thought and, for example, coming up with alternative concepts.

Box 4.6 Examples of meta-questions and test objectives

Energy consumption: Do you now take greater care than before with your energy consumption because energy prices are so high?

Test objective:

What do people mean when they say that they do or do not take greater care with their energy consumption?

Meta-questions:

If someone says yes: ask for examples of how, since when, and why they do so.

If someone says no, ask whether they mean ‘no, I don’t take great care with my energy consumption’ or ‘no, I do take great care, but for other reasons / I always have taken great care with my energy consumption’.

Recycled paper How often do you use recycled paper?

Test objective:

In the field a fairly common (4%) response to this question is ‘don’t know’. What does this mean: that people do not know what recycled paper is, or that they are not always sure whether the paper they use is recycled, or something else?

Meta-questions:

If someone answers: can you repeat this question in your own words? How did you interpret this question?

If someone refrains from answering: what did you find unclear in the question?

- Recruitment and selection of test respondents

Most cognitive interviews are conducted with for this purpose recruited test respondents. Sufficient diversity of test respondents is necessary in order to guarantee that they may cover most of the possible problems in a questionnaire. This may mean, for example, recruiting a disproportionately large number of self-employed, low-skilled or ill people, where these are the target populations expected to have problems with the questionnaire. Leaving subject-specific matters aside, it is advisable when testing social surveys to arrange for a good distribution of gender, age, educational level and employment status. When testing business surveys a good distribution of size category and industry is desirable.

The required number of test respondents depends on an assessment of the risks of problems with the draft questionnaire and the number of different routes to be tested. The number of test respondents recruited would normally be between 20 and 30.

- Role of the cognitive interviewer

  It is important for interviewers of cognitive interviews to adopt a neutral stance. In particular someone testing their own questions will often tend to correct respondents who make mistakes, or defend the way the question is phrased. This should not happen, since it will bias the outcomes of the cognitive interview. In business surveys, where there is often a long-term relationship between the respondent and Statistics Netherlands, it may be worthwhile explaining to respondents the meaning of questions. This explanation should always be given after the actual cognitive interview has ended.

  Test respondents in cognitive interviews must be given thorough instructions about their task, emphasizing the need to test the questionnaire rather than the respondent. Respondents will often be irritated if they misunderstand a concept, or answer a question wrongly. If possible they should not even be told when they make mistakes.

- Execution

  Cognitive interviews are often held behind closed doors, with the discussions recorded on video or audio media. The recordings are used for further analysis and possibly as illustrations in a presentation of the test results. Experience has shown that a one-minute film of a test respondent completely misinterpreting a question can be more convincing than a twenty-page memo.

  For business surveys cognitive interviews are usually held on the respondent’s or company’s site, because there access to records will be available. For electronic questionnaires it may be useful to capture screen shots. (See also Section 4.8 Observing respondents and interviewers).

  Cognitive interviews are executed in accordance with protocols, which describe the objectives of the test, questions where difficulties are expected, and instructions for interviewers. It is advisable to use an answer or report template in which interviewers can record any relevant reaction of the test respondent to each question.

Advantages and disadvantages: Cognitive interviews provide qualitative information collected from a small group of respondents. Interpretation and generalization of the results can sometimes be difficult. Moreover, recruiting respondents can be a labour intensive exercise. Other aspects that are viewed as unfavourable are the lenghty test period and the verbal skills required of test respondents. On the other hand, each cognitive interview yields a wealth of information about the question-answer process, which allows to identify errors in good time. In our experience no other test method provides as much insight into the causes of measurement error as the cognitive interview. We therefore endorse the Eurostat recommendation at least to test every questionnaire with several test respondents drawn from the envisaged target population.

4.7 Vignettes and case-based questionnaire answering

Brief description: Vignettes are scenarios to which test respondents respond, thereby giving extensive evidence of their understanding of a concept. Test respondents may also be asked to fill in an entire questionnaire with reference to a practical example, which is referred to as a case.

a) Vignettes

Purpose: Respondents' comprehension of complex concepts in surveys may be unclear. It can be equally unclear whether all respondents have the same understanding. In order to find this out numerous vignettes can be presented to a series of test respondents.

Phase: Usually at an early stage of questionnaire development, often in combination with a focus group.

Extensive description: A vignette research starts with a difficult concept to which respondents attach maybe a wide range of meanings, attitudes or emotions. Concepts of this kind should normally be avoided in surveys, but these are sometimes inevitable. An example is the concept of 'changing job' that is explored in the Labour Force Survey (LFS, in Dutch EBB). In order to find out what respondents understand by 'job' and 'changing job', up to a few dozen situations may be described in detail. All situations must be realistic examples of something that falls just inside or outside the definition of changing job. The test has been performed with test respondents, but was carried out at a much earlier stage with Statistics Netherlands employees. Many different definitions were encountered in both cases. After test persons have answered the questions, e.g. with 'change' or 'no change', a focus group can be held and the outcomes compared. Individual follow-up interviews could also be held, or respondents could be asked to think out loud when assessing the vignettes.

Advantages and disadvantages: A vignette survey is an excellent means of rendering a target population's perceptions of concepts and constituent aspects. The output can provide a basis for a description to be used in the questionnaire, or support a decision to divide the concept into diverse questions, or conversely not to use it at all. A disadvantage is the strong dependence on the quality of the vignettes, that is how realistic they are and how well respondents can imagine the described situations.

b) Case-based answering

Purpose: Sometimes a questionnaire has to be tested which is meant only for special subpopulations or circumstances. It is not always easy to find test respondents who exactly match the specific requirements. Certain questions or questionnaires can feasibly be tested with reference to a real or fictional case and using test respondents who, while not meeting the set requirements perfectly, are able to act as proxies. Field staff with substantial experience in the field are a good example. Cases or practical examples of this kind can also be used to obtain satisfactorily comparable results about how well a questionnaire performs. For instance, two alternative questionnaires could usefully be filled in on the basis of the same cases.

Phase: Usually in combination with focus groups in the test phase.

Extensive description: Test respondents in this method are given a brief description, case or situation (vignette) (for an example, see Box 4.7). They are requested to base their responses on this information. Researchers may ask test respondents to think out loud while answering, in order to clarify the answering behaviour. The same may happen in a focus group, by requesting the group to complete the questionnaire with reference to one or more prescribed situations. Certain decisions can then be discussed in the group.

The results of this method are an indication of possible problems with questions oriented to specific target populations.

A good case is a realistic scenario comprising all the information needed to perform an answering task. Preferably place the information in a natural sequence (and not necessarily in the same sequence as the information is requested in the questionnaire). If the information has to be retrieved from existing documents (e.g. an insurance policy or payroll records), include the necessary documents in the case information. The more realistic the answering (and retrieval) task, the more valuable the test.

Box 4.7 Example of a case

One of the participants in the Intelligible Forms workshop given by Statistics Netherlands in cooperation with the Ministry of the Interior was asked to refer to the case below and complete the *'Residence/departure/relocation registration form'* of the municipality of Heerlen.

Case: Complete the form in the name of 19-year-old Maurice van der Wal's father:

Maurice will be leaving his parent's home in Heerlen for college in Eindhoven, where he and his best friend Tim will be renting a three-room flat. He has authorized his father, Jan van der Wal, to complete the form.

Subsequently observing and discussing this participant's style of completing the form makes it possible to identify any problems with the form in how the questions are phrased, or the answer options.

Advantages and disadvantages: A test that uses practical examples enables a questionnaire intended for a specific target population to be tested in little time and with few resources. Questions for which it is almost impossible to find test respondents can be tested in this way. Furthermore an attractive aspect of these practical examples is that the input is standardized, allowing systematic comparison of multiple questionnaire versions. There are two major disadvantages to a case approach. The first is that the technique demands substantial cognitive skills and creativity of the test respondent in processing and understanding the information in the given example. The second is that the respondents act as proxies and as such are never really representative of the target population.

4.8 Observing respondents and interviewers

Brief description: When observing respondents or interviewers, certain forms of behaviour are registered systematically. For instance, when playing back several past interviews it is noted which questions caused problems, such as the interviewer not reading the question texts literally, or the respondent interrupting the interviewer, or being unable to choose an answer option immediately. Survey questions that often cause this kind of behaviour should probably be revised.

Purpose: By observing and then analysing the question-answer process it is possible to identify where respondents and interviewers encounter problems in asking or answering questions. It is a way of investigating the effectiveness of a questionnaire.

Phase: Observations may be made once a draft questionnaire is available, usually after the questionnaire design phase and in the test phase. When revising an existing survey the method can be used to investigate how a current questionnaire is working in the field.

Extensive description: There are many ways of observing. Some are entirely inconspicuous, e.g. as in recording CATI interviews, or producing an audit trail, or very conspicuous, an extreme example of which is recording a respondent's eye movements using special head-mounted equipment.

Several ways of observing are described briefly below.

- Behavioural observations. These involve observing a respondent's or interviewer's behaviour. Attention is given to body language and spoken language, since both are important.
- Conversation analysis. This involves analysing only the dialogue between the respondent and the interviewer. Body language is disregarded, but paralinguistic elements (such as sighing or a raised voice) may be coded. Attention is given to the respondent's requests for clarification of a question, or periods of silence before answering, or whether an interviewer alters a prescribed question text. Box 4.8 gives a specimen transcription of a telephone interview. For additional information, see: Mazeland (1992) or Psathas (1995).
- Screen capture. This refers to preserving a copy of the display screen that is visible to a respondent, recording every mouse movement, and everything that is entered. This method allows to investigate the usability of a web questionnaire and to study the answering behaviour. E.g. is a respondent able to find the right button to browse through the questionnaire, or do respondents notice the button for additional information?

Box 4.8 Example of conversation analysis of a question-answer sequence from the 2002 Consumers Confidence Survey

Transcription:			Coding scheme:
96.	I	What do you expect to happen in the next twelve months: will the economic situation in the Netherlands improve, deteriorate, or stay the same?	All the sentences are numbered. I = texts of the interviewer. R = texts of the respondent. [ ] = The square brackets show where the interviewer and respondent spoke at the same time. (0.0) = Figures in parentheses show the length of a silence. A silence too short to express in seconds, but noticeable nonetheless, is indicated with a point only. : = A colon after a sound indicates that it persisted for longer than normal.
97.	R	(.) Mm: I think it'll get better.	
98.	I	And what would you say, a bit better or much better?	
99.	R	I think a bit better.	
100.	I	The prices. Do you think prices have risen in the past twelve months, gone down, or stayed about the same?	
101.	R	Uh (.) I think they've gone up a bit.	
102.	I	(2.0) And should I enter moderate or weak as your answer?	
103.	R	The rise, do you [mean?]	
104.	I	[Yes.]	
105.	R	Yes, I'd say weak.	

- Observation in a cognitive interview. Hak & Muller (2001) describe the mix of observation and interview as the Three-Step Test Interview. In Step 1 the test person performs a task (such as completing a questionnaire) while thinking out loud. In Step 2 the researcher and test person reconstruct what happened as completely as possible, with the aim of obtaining the clearest possible picture to augment the observation made in Step 1. Finally, in Step 3, a qualitative interview is held to explore supplementary information and opinions.
- Audit trails. Audit trails are sometimes known as log files. Audit trails record all actions taken by a respondent on the computer: the timing of every mouse click and every keystroke. Like screen capture, an audit trail provides a picture of a respondent's progress through a web questionnaire. Data of this kind about the survey process are known as para data. For additional information see Box 4.9.

Statistics Netherlands has software available for coding and analysing behaviour recorded on video or audio media: Interact.

Box 4.9 Explanation of audit trails

A respondent's actions while completing a questionnaire can be captured in an audit trail along with the date and time.

The actions people perform and the amount of time they spend on them can be derived from audit trails. They can also reveal the route people take through a questionnaire, which questions were changed in the process of answering, the number of attempts needed to answer a question and when the help function was used.

Audit trail applications are still in development. The extremely diverse opportunities include identifying difficult questions, questions that are completed as a block, problems with the questionnaire program, and types of user (Morren & Snijkers, 2008).

Advantages and disadvantages: A realistic picture of the question-answer process can emerge if observation is inconspicuous. It may yield data from real situations in the field. Observation – both in the field and in the laboratory – yields rich and realistic data about such things as answering behaviour, the usability of a questionnaire and interviewer-respondent interactions. The greatest disadvantage of behavioural observations is that both data collection and analysis are time consuming.

4.9 Analysis of qualitative survey material

The information collected in focus groups, cognitive interviews and observations is usually qualitative in nature. It is easy to lose the overall view with rich material of this kind, and it is therefore advisable to process the data in a systematic way.

A practical approach is as follows. Refer to the test protocol or topic list and make a table with the investigated survey questions together with any points for attention and meta-questions in the column headings, while labelling the rows with the numbers of the test persons or focus groups. Then watch or listen to the recordings and for each test person or focus group note the findings in the appropriate rows and columns. Any problems with the questionnaire that emerge must be comprehensively described, preferably giving literal quotes, and adding initial ideas about the cause and solution. After documenting all interviews or discussions in this way, it is useful to discuss the raw analysis scheme with several researchers. Preliminary hypotheses about the questionnaire are then formulated. E.g. 'There are too many response categories for Question x' or 'Question y cannot be answered by people whose children have left home'. Some problems and solutions that emerge are obvious, e.g. if a relevant answer option is missing. Other problems demand careful analysis of the material, testing the preliminary hypotheses against the raw analysis scheme or the original recordings.

The above method is particularly suitable for systematically investigating the suspected risks that were in mind when formulating the meta-questions, thereby creating a view of the nature and scale of these problems. Urgent and less urgent changes can then be proposed.

On the other hand it is important when watching or listening to be open to unsuspected problems. However experienced a questionnaire developer or tester may be, entirely unexpected things may be revealed. This is why recordings are so necessary, and the presence of a detached observer – at least in some interviews – is also desirable. The group discussions mentioned above can also help in the analysis by revealing unsuspected problems. It may then be advisable to study the recordings again to ascertain whether the problems are incidental or structural.

Qualitative research– suggestions to the contrary notwithstanding – deserves the same systematic and thorough analysis as quantitative research. It also demands equally systematic and thorough data collection. For an overview, see Denzin & Lincoln (2003).

4.10 Technical testing

Brief description: Technical (also known as functional) testing involves checking whether a computer-assisted questionnaire was built in accordance with specifications.

Purpose: The purpose of technical testing is to detect errors in questionnaire construction. The tests check such things as whether all routes in the questionnaire flow properly, and that all practical situations have been taken into account.

Phase: These tests are always performed after all other tests and before the data collection phase.

Extensive description: The questionnaire is tested by means of the 'Test specification technique of the process cycle' (Pol, Teunissen & Veenendaal 2000). This technique first identifies the routings (decision points) in the questionnaire by assigning a number to every exit path from a decision point in the questionnaire. The test situations are then defined. This lists all combinations of paths around a decision point. Finally all the test situations are recombined into a complete path from the start to the end of the questionnaire. As many complete paths are defined as are needed to cover all the test situations in the test. These paths can be translated into protocols, writing out each answer to be given for each question. Technical tests of this kind are executed prior to the data collection phase. The main point is to test the process or system to be used in collecting the data. Another aspect is testing the technical test specifications of web surveys. Some examples are automatic calculations, error messages, and submitting the completed web questionnaire.

Advantages and disadvantages: This technique yields unambiguous test results: an error is either found or not found. The rectification of questionnaire errors and then adjusting the test design demands effective coordination of the various people involved, such as data process experts, researchers and questionnaire developers.

4.11 Experiments

Brief description: Experimental questionnaire research usually involves presenting a random sample from the research population with alternative phrasings of questions or questionnaires. Conditions are manipulated in experiments (e.g. the versions of a question). Respondents are randomly assigned to these conditions.

Purpose: Experiments can test and check multiple variants of a questionnaire or multiple modes in order to detect measurement errors. A motive for conducting an experiment would be to choose between two alternative phrasings of a question where no decision emerged in an earlier discussion or pre-test. Experiments are also often used for investigating possible problems with time series because of rephrasing, or applying a different mode.

Phase: Experiments can be conducted as part of a pilot in the test phase or embedded in ongoing surveys, i.e. in the data collection phase.

Extensive description: Experiments may involve individual questions, a set of questions, or an entire questionnaire. A condition is that the sample size for an experiment must be large enough to allow significant differences to be measured. It is also necessary for the design to permit random allocation of questionnaire variants, so that any differences can be attributed to the questionnaire rather than the composition of the sample. See also Van den Brakel (2008) regarding experiments held in the context of normal data collection. For a checklist for setting up an experiment see Box 4.10.

Advantages and disadvantages: Setting up and conducting an experiment is an intensive and relatively expensive and time consuming way of testing. However, the output does deliver reliable and quantifiable conclusions about question variants or modes.

Box 4.10 Experiment setup checklist

- describe the problem;
- determine the variants of the questions and questionnaires;
- formulate hypotheses that can be tested;
- define design and size of the experiment;
- carry out the experiment;
- test the hypotheses and interpret the results;
- make decisions and implement.

(Statistics Sweden, 2004, Page. 65)

4.12 Process data analysis

Brief description: Much data about the data collection process can be recorded when the questionnaire is finally in the field. These data are known as para data. Analysing these data can help in understanding how a questionnaire performs.

Purpose: Besides various test methods prior to questionnaire implementation, it is also possible after implementation when a questionnaire is in the field for data collection, to investigate whether the data collection instrument meets the chosen quality objectives.

Phase: It is logical for the process data to be collected in the data collection phase, and analysed both during data collection (e.g. after the first ten days of a fieldwork period or the first 100 respondents) and afterwards.

Extensive description: Process data or para data are defined as the data recorded about the data collection process. Examples include information about how often a reminder was issued to a respondent, at what time an interview was administered, or e.g. an audit trail of precisely how a computer questionnaire was used. These data might be an unintentional data collection by-product, or collected specifically for evaluation purposes. We give below a number of frequently used methods of using process data for evaluating questionnaires.

- Analysis of the level and duration of response: figures about the unit response and the time between sending a questionnaire and receiving completed or partly completed questionnaires can provide insight into which questionnaires, or which target populations, are giving problems.
- Quantitative or qualitative analysis of respondents' comments and questions recorded in the Customer Relations Management (CRM) system can provide insight into which questionnaires create most problems, and the nature of the problems. See Box 4.11 for an example. The above applies to business questionnaires. What is important for social surveys is respondents' comments recorded by interviewers, or interviewers' questions received by Statistics Netherlands.
- Meta-questions in the issued questionnaire: this method involves including meta-questions in the questionnaire to be presented to respondents alongside certain questions, or at the end of the questionnaire. The questions could be about the response burden and perceptions of the questionnaire. This provides a way of measuring the quality of the question-answer process and identifying awkward or sensitive questions. Statistics Netherlands uses this method infrequently, but many market research bureaus systematically ask a few questions at the end of an interview, in part to identify themes that are popular or unpopular. A technique of putting these questions in a separate mini-interview is known as Response Analysis Survey (RAS), as described below.
- A Response Analysis Survey (RAS) is a standard follow-up interview with respondents and non-respondents. RAS is used mainly for a) interviewer monitoring, and b) nonresponse analysis. Nonresponse analysis involves asking non-respondents a few key questions in order to gain insight into their common attributes. The Statistics Netherlands annual customer satisfaction survey among respondents and non-respondents of certain surveys can also be viewed as a form of RAS.

- Brief questionnaire about process data for interviewers. Interviewers too can give much information about the process, e.g. whether anyone else was present during the interview with the respondent, or whether any questions came over as too intrusive. Questions of this kind are presented to interviewers in a separate list.

Advantages and disadvantages: These methods yield data about the actual data collection process that usually lend themselves to quantitative analysis. Reliable statements can then be made about where quality problems are occurring and how serious they are. There are two major disadvantages of analysing process data. The first is that problems can be detected only after entering the field. It would clearly be better to prevent these problems. The second is that these methods usually shed little light on the causes and possible solutions to the problems. Any insights based on process data from a revision round are therefore best used as input for further qualitative research.

Box 4.11 Example of CRM data analysis

Results of coding respondents' comments about the 2003 Production Statistics (Giesen & Van Putten, 2004).

	Main code	Number	Percentage
1	Comments about specific questions / question blocks	291	13.0
2	Comments about lack of clarity of questionnaire in general	26	1.2
3	Comments about compatibility between questionnaire and administrative system	610	27.3
4	Comments about response burden	665	29.8
5	General comment about answer quality	179	8.0
6	Other comments	461	20.7
	Total	2232	100

4.13 Survey data analysis

Brief description: Questionnaires determine to a great extent the quality of the data. Problems with the quality of the collected data may therefore be an indication that the questionnaire is lacking in quality.

Purpose: The purpose of analysing survey data as a method of evaluating questionnaires is usually to assess the quality of the collected data and to investigate the part caused by errors in the questionnaire.

Extensive description: There are various ways of assessing survey data quality. The aspects usually investigated in the questionnaire evaluation are item nonresponse, data plausibility, response patterns in answer scales, and the distribution of response options. An extensive discussion of this topic is outside the scope of this part of the Methods Series. Useful synopses of quality measures for survey data are given in Lyberg et al. (1997).

Advantages and disadvantages: The advantages and disadvantages of survey data analysis are comparable with those of process data analysis. An additional disadvantage has to do with the many aspects of the data collection process, besides the questionnaire itself, that have an impact on data quality (e.g. reminders, support to field staff, and the data collection mode). It is not always possible unambiguously to determine the questionnaire's exact direct contribution to quality problems in the survey data.

5. Conclusion

Statistics Netherlands tests most new and revised questionnaires. Testing is performed in a variety of ways in different departments. There exists also a group of survey methodologists who specialize in testing questionnaires: the Questionnaire Laboratory (Vragenlab).

The Lab advises questionnaire developers after testing new and revised questionnaires. The Lab uses all the test methods discussed in Chapter 4. The Lab is open to questions from questionnaire developers about invitation letters, data collection methods, question phrasing, questionnaire usability and applicability, and for pilot surveys and experiments.

A single test programme will rarely use all the methods for evaluating questionnaires, which would usually exceed the available time and resources. There is typically little time left for testing a questionnaire because of delay in earlier development phases. A selection will have to be made based on objectives and resources. The following minimum package for a test and evaluation programme is derived from Dehue (1997) and Snijkers (2002):

Table 5.1. Minimum package for test and evaluation programme

Phase	Questionnaire development	Methods
1 & 2	Conceptualization and survey & questionnaire design	▪ Document and literature review, informal test; ▪ expert review; ▪ if the above is less than convincing: demonstration with focus group and/or cognitive interviews.
3	Testing	▪ Observation of several standard interviews with test respondents, combined with overall evaluation questions about questionnaire content and procedure.
5 & 6	Data collection, and process monitoring and evaluation	▪ Observation of several interviews in the field, combined with evaluation questions for respondent and/or interviewer; ▪ follow-up interviews with respondents; ▪ evaluative focus groups with interviewers; ▪ item nonresponse analysis; ▪ data analysis.

Experience has shown that even this minimum package is not always feasible. Our advice is always to test every questionnaire with several test persons from the target population, and additionally to use as many different methods as possible. There is good reason to have confidence in conclusions that emerge consistently from a variety of methods. It is also most important to feed back results from the tests and evaluations to researchers and questionnaire developers, to facilitate adjustment and improvement. As stated above in Chapter 2 (see also Figure 2.1), questionnaire development is an iterative process, where information from each phase in the process can serve as input for improving a draft questionnaire or updating an existing one.

6. References

Beukenhorst, D., Dam, G. van, Göttgens, R., Snijkers, G. and Verbruggen, M. (2005), *CBS brede waarneemstrategie (Statistics Netherlands data collection strategy).* Internal memo, Statistics Netherlands, Heerlen.

Biemer, P. and Lyberg, L. (2003), *Introduction to Survey Quality*. Wiley, New York.

Billiet, J. (2006), The standard questionnaires. In: Billiet, J. and Waege, H. (Eds.). *Een samenleving onderzocht (A society surveyed).* De Boeck, Antwerp, pp. 223-283.

Brancato, G., Macchia, S., Murgia, M., Signore, M., Simeoni, G., Blanke, K., Körner, T., Nimmergut, A., Lima, P., Paulino, R. and Hoffmeyer-Zlotnik, J. (2006), *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System.* European Commission Grant Agreement 2004103000002.

Bureau Taal (2006), *Schrijfwijzer Begrijpelijke Formulieren (A guide for writing comprehensible forms)*. SDU. www.begrijpelijkeformulieren.nl

Campanelli, P. (2008), Testing survey questions. In: De Leeuw, E., Hox, J. and Dillman D. (Eds.). *International Handbook of Survey Methodology*. Lawrence Erlbaum, New York, pp 176-200.

City of The Hague. Website consulted on 27 March 2009. *Helder Haags Woordenboek (Transparent dictionary for The Hague).* http://www.denhaag.nl/smartsite.html?id=55844.

Couper, M. (2008), *Designing effective web surveys.* Cambridge University Press, Cambridge.

Dehue, F. (1997), *Standaardprotocollen voor de ontwikkeling en de test van dataverzamelingsprocedures (Standard protocols for the development and testing of data collection procedures)*. Internal report, Statistics Netherlands, Heerlen,.

De Leeuw, Hox, J. and Dillman, D. (Eds.) (2008), *International Handbook of Survey Methodology*. Lawrence Erlbaum, New York.

Denzin, N., Lincoln, Y. (Eds.) (2003), *Collecting and interpreting qualitative materials*. Sage, Thousands Oaks.

Dillman, D., Smyth, D. & Christian, J. (2009), *Internet, Mail and Mixed Mode Surveys. The Tailored Design Method.* Wiley, New York.

Dillman, D. and Redline, C. (2004), Testing paper self-administered questionnaires: cognitive interview and field test compared. In: Presser, S., Rotgheb, J., Couper, M., Lessler, J., Martin, E., Martin, J. and Singer, E. (Eds). *Methods for Testing and Evaluating Survey Questionnaires.* Wiley, New York, pp 299-317.

Eurostat (2011), European statistics code of practice for the national and community statistical authorities. Adopted by the European Statistical System Committee 28th September 2011 http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-11-955/EN/KS-32-11-955-EN.PDF

Fowler, F. (1995), Improving survey questions: design and evaluation. *Applied Social Research Methods Series*, vol. 38, Sage Publications.

Fowler, F., Roman, A. & Di, Z. (1998), Mode effects in a survey of Medicare prostate surgery patients. *Public Opinion Quarterly* 62, 29-46.

Fowler, F. and Cosenza, C. (2008), Writing effective survey questions. In: De Leeuw, E., Hox, J. and Dillman D. (Eds.), *International Handbook of Survey Methodology*. Lawrence Erlbaum, New York, pp 136-160.

Giesen, D. (2004), *Evaluatie vragenlijsten productiestatistieken: Resultaten en aanbevelingen (Evaluation of questionnaires production statistics: results and recommendations).* Internal report, Statistics Netherlands, Heerlen.

Giesen, D. and Putten, J. van (2004), Analyse van opmerkingen van berichtgevers in Logiquest (Analysis of Logiquest respondent comments). In PS project team evaluation *Resultaten Inventarisatiefase Evaluatie Vragenlijsten Productiestatistiek (Production statistics questionnaire evaluation inventory phase output)*. Internal report, Statistics Netherlands, Heerlen.

Grice, H.P. (1975), Logic and Conversation. In Cole, P. & Morgan, J.L. (Eds.), *Speech Acts.* Academic Press, New York, pp 41-58.

Hak, T. and Muller, J. (2001), De toolkit van de kwalitatieve marktonderzoeker. Het Drie-Stappen Test-Interview (The qualitative market researcher's toolkit. The three-step test interview). *Onderzoek*, 18 (12), pp 8-9.

Houtkoop-Steenstra, H. (2000), *Interaction and the standardized survey interview: the living questionnaire.* Cambridge University Press.

Jansen, C., Steehouder, M., Edens, K., Mulder, J., Pander Maat, H., Slot, P. (1989), *Formulierenwijzer. Handboek formulieren redigeren (Forms guide. Form editing handbook).* SDU, The Hague.

Krosnick, J. (1991), Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5, 213-236.

Krosnick, J., Holbrook, A., Berent, M., Carson, R., Hanemann, W., Kopp, R., Mitchel, R., Presser, S., Ruud, P., Smith, V., Moody, W., Green, M. and Conaway, M. (2002), The impact of "no opinion" response options on data quality. *Public Opinion Quarterly* 66, 371-404.

Krueger, R. A. (1994), *Focus groups: A practical guide for applied research*. Sage Publications, London.

Lyberg, L., Biemer, P., Collins, M. de Leeuw, E., Dippo, C., Schwarz, N. and Dennis, T. (Eds.) (1997), *Survey Measurement and Process Quality*. Wiley, New York.

Maynard, D., Houtkoop-Steenstra, H., Cate Schaeffer, N. and Zouwen, J. van der (Eds.) (2002), *Standardization and Tacit Knowledge.* Wiley, New York.

Mazeland, H. (1992), *Vraag/antwoord sequenties (Question-answer sequences).* Stichting Neerlandistiek VU, Amsterdam.

Morren, M. and Snijkers, G. (2008), *Utility of audit trails in the annual structural business survey.* Internal report, Statistics Netherlands, Heerlen.

Pol, M., Teunissen, R. and Veenendaal, E. van (2000), *Testen volgens TMAP (Testing in accordance with TMAP)*. Tutein Nilthenius, ‘s Hertogenbosch (par. 15.11 Procescyclustest).

Psathas, G. (1995), *Conversation Analysis: the study of talk-in-interaction.* Sage Publications Inc., Thousand Oaks.

Redline, C. and Dillman, D. (2002), The influence of alternative visual designs on respondents’ performance with branching instructions in self-administered questionnaires. In: R. Groves, Dillman, D., Eltinge, J., and R. Little (Eds.), *Survey Nonresponse.* Wiley, New York.

Schwarz, N., Knäuper, B., Hippler, H-J., Noelle-Neuman, E. & Clark, F. (1991), Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly* 55, 618-630.

Snijkers, G. (2002), *Cognitive laboratory experiences: on pre-testing computerised questionnaires and data quality*. Thesis: Statistics Netherlands/UU Utrecht.

Statistics Sweden. (2004), *Design your questions right*. Sweden.

Sudman, S and Bradburn, N M (1983). *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass , San Francisco, CA.

Toepoel, V., Vis, C., Das, M. and Soest, A. van (2009), Design of Web Questionnaires. An information-processing perspective for the effect of response categories. *Sociological Methods & Research* 37, 371-392.

Tourangeau, R., Rips, L.J. and Rasinski, K. (2000), *The Psychology of Survey Response*. Cambridge University Press, New York.

Van den Brakel, J.A. (2008), Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey. *Journal of the Royal Statistical Society* (Series A) 171 (3), 581-613.

Vis-Visschers, R. (2002), *Conversatie Analyse en het testen van survey-vragen (Conversation analysis and survey question testing).* Internal report, Statistics Netherlands, Heerlen.

Vis-Visschers R. (2007), *Dutch Testing Phase Feedback of the Questionnaire on International Sourcing*. Internal report, Statistics Netherlands, Heerlen.

Vis-Visschers, R., Arends-Tóth, J., Giesen, D. and Meertens, V. (2008), *Het aanbieden van weet niet en toelichtingen in een webvragenlijst (Providing don't know and explanations in a web questionnaire).* Internal report, Statistics Netherlands, Heerlen.

Vis-Visschers, R., Morren, M., Meertens, V. and Arends-Tóth, J. (2008), *CBS cognitive test of the health care module and the health determinants module of the European health interview survey*. Internal report, Statistics Netherlands, Heerlen.

Vosmer, M. and Engelen, F. (2008), Het herzien van de vragenlijst van het NKO 2006 (Revision of the 2006 national electoral opinion poll questionnaire). In H. Schmeets (Ed.). *Het Nationaal Kiezersonderzoek 2006. Opzet, uitvoering en resultaten (The 2006 national electoral opinion poll. Setup, execution and results).* Statistics Netherlands, Voorburg/Heerlen, 55-61.

Willeboordse, A. (2008), Methodenreeks: *Inleiding in het statistisch proces en de Methodenreeks (Introduction to the statistical process and the Methods Series).* Statistics Netherlands, Voorburg.

Willimack, D. and Nichols, E. (2001), Building an Alternative Response Process Model for Business Surveys. *U.S. Proceedings of the Annual Meeting of the American Statistical Association*, Alexandria, V.A.: American Statistical Association.

Willis, G. (2005), *Cognitive interviewing: a tool for improving questionnaire design.* Sage Publications, London.

Appendix 1. Language proficiency levels

Reference levels set out in the Common European Framework of Reference for Languages (CEFR)
http://www.coe.int/t/DG4/Portfolio/?L=E&M=/main_pages/levels.html

		A1	A2	B1	B2	C1	C2
UNDERST	**Listening**	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided. I have some time to get familiar with the accent.
ANDING	**Reading**	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

SPEAKING	**Spoken Interaction**	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.
WRITING	**Writing**	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.	I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

Appendix 2. Questionnaire development checklist

How to use this checklist?

This checklist can be used to ensure you have thought of everything and that during the development process you have made well-considered decisions. It is not compulsory that each item is scored with a 'yes'; there may be good reasons for scoring 'no', or 'more or less'.

If there is no good reason or well-considered basis for answering 'no' or 'more or less', then it is advised to consider modifying the questionnaire.

The checklist is an adaptation of the checklist in the handbook *Formulierenwijzer (Guide for making forms. In Dutch)*(Jansen et al., 1989)

		Yes	More or less	No	Not applicable
	A. In general				
1.	Does the completed questionnaire answer the research questions?	□	□	□	□
2.	Is the requested information restricted to what is necessary? ➢ Do not ask for information that is already known. ➢ Never include a question 'because it might be interesting'.	□	□	□	□
3.	Is sufficient general information provided about the questionnaire's purpose?	□	□	□	□
4.	Does the advance letter or introduction to the questionnaire provide sufficient information about what is expected of the respondent?	□	□	□	□
5.	Does the questionnaire have a clear and inviting title? ➢ Questionnaires should have a title that is meaningful to respondents.	□	□	□	□
6.	Is the requested data easy to provide by the respondent?	□	□	□	□
7.	Has the questionnaire been tested with test respondents from the target population?	□	□	□	□

		Yes	More or less	No	Not applicable
	B. Phrasing of questions				
1.	Are all questions expressed in the form of a question? ➢ End each question with a question mark.	□	□	□	□
2.	Are all questions singular? ➢ Ask only one question at a time. Words such as ‘and’, ’or’, ‘because’ and ‘because of’ are often indicators that there is more than one question hidden in the question text.	□	□	□	□
3.	Are the questions unambiguous? ➢ Provide definitions of terms that could be interpreted differently or are unclear.	□	□	□	□
4.	Is it always clear to whom, what and which period the questions refer? ➢ If necessary provide reference periods. ➢ Choose a reference period that is relevant to the importance of the event – the less important an event, the shorter the reference period.	□	□	□	□
5.	Are the questions free of assumptions? ➢ Do not say: ‘As you know... ’.	□	□	□	□
6.	Are double negatives avoided?	□	□	□	□
7.	Are the questions and sentences brief?	□	□	□	□
8.	Are the questions written in plain language? ➢ Choose short and everyday words. ➢ Be consistent in vocabulary. ➢ Do not use jargon or abstract terms. ➢ Phrase texts actively and positively.	□	□	□	□
9.	Are the questions easy to read out if the survey is interviewer administered? ➢ In CAPI and CATI surveys interviewer friendliness is also important.	□	□	□	□
10.	Are the spelling and punctuation used correctly?	□	□	□	□

		Yes	More or less	No	Not applicable
	C. Response options				
1.	Does the question make clear what kind of answer is expected? ➢ The response options should fit the question.	□	□	□	□
2.	Are all possible response categories provided? ➢ If necessary, include a catch-all category: “Other, that is: …”. ➢ Always put the catch-all category at the end of the list. ➢ When presenting a list of possible options, put the most likely answer first. ➢ Explicitly include the response options ‘don’t know’ or ‘refuse’ when they are relevant responses.	□	□	□	□
3.	Are the response categories mutually exclusive? ➢ Avoid overlap between the response options.	□	□	□	□
4.	Is the response scale balanced? ➢ When using scales, make sure that they are balanced, e.g. as many positive as negative options.	□	□	□	□
5.	Is the meaning of the response options clear? ➢ Are the response options unambiguous?	□	□	□	□

		Yes	More or less	No	Not applicable
	D. Explanations and instructions				
1	Are explanations and instructions given only where necessary? ➢ The more text, the less likely that people will read it.	□	□	□	□
2.	Are the explanations and instructions located as near to the question as possible? ➢ Try to include instructions as much as possible in the question texts; keeping the previous point in mind. ➢ Do not “hide” instructions behind clickable buttons, or in separate appendices of the questionnaire.	□	□	□	□

3.	Are the explanations and instructions plain and clear? ➢ Keep sentences short. ➢ Beware of words such as 'excluding' and 'not'; respondents tend to overlook this.	□ □ □ □
4.	Are hints provided as to where the relevant information can be found, if respondents have to look up information in order to be able to answer a question? ➢ Look up tasks are burdensome, try to make it as easy as possible for the respondent.	□ □ □ □

		Yes / More or less / No / Not applicable
	E. Structure of the questionnaire	
1.	Is the numbering of questions, blocks and pages clear and correct?	□ □ □ □
2.	Is the route through the questionnaire clear? ➢ An interviewer or respondent should always know the current position within the questionnaire. ➢ The routing instructions should be unambiguous and clear. Take the respondent as much as possible by the hand.	□ □ □ □
3.	Are the questions clustered into recognizable themes (blocks)? ➢ Each block deals with one defined theme or task. People work from task to task: structure the work in clearly defined subtasks. ➢ Give each block a clear separate title.	□ □ □ □
4.	Is the sequence of blocks logical for the respondent and also favourable for the response? ➢ Preferably ask about background characteristics last. ➢ Do not start a questionnaire with sensitive questions, nor with difficult questions.	□ □ □ □

		Yes	More or less	No	Not applicable
	F. Format / visual design				
1.	Is the chosen design functional? ➢ Any variation in layout, every symbol, etc. must have a function. ➢ Any symbol used in the questionnaire should be clear and explained if necessary.	□	□	□	□
2.	Is the chosen design applied consistently throughout the questionnaire? ➢ Layout, symbols, line spacing etc. must have the same meaning throughout the questionnaire. ➢ Always put the same kind of information in the same place and in the same style.	□	□	□	□
3.	Is the use of colours conscientious? ➢ Use contrasting colours as a complementary design feature. ➢ The use of colours should also be effectively for people with vision impairment.	□	□	□	□
4.	Does the design (visually) support the structure of the questionnaire? ➢ The design can make clear which items in a questionnaire belong together. ➢ Make sure that it is always clear which response box belongs to which question. ➢ Make sure that there is enough space for the respondent to write down his answer.	□	□	□	□
5.	Is any information which is not intended for the respondent (e.g. coding keys, respondent numbers) formatted less prominently? ➢ What respondents do not need to know, they do not need to see; it could also distract or confuse them.	□	□	□	□

		Yes / More or less / No / Not applicable
	G. Specific points for Internet questionnaires	
1.	Is it clear where and how the questionnaire can be downloaded, opened and sent? ➢ Make sure the web address is short and clear. Typing errors are easily made ➢ Keep the web page clear and surveyable. ➢ Does the respondent receive a confirmation of receipt after sending in the electronic questionnaire?	□ □ □ □
2.	Have the questionnaire programmers considered downward compatibility? ➢ Most people do not have the most up-to-date soft- or hardware; you want the electronic questionnaire to work for most browsers and software.	□ □ □ □
3.	Does the respondent always know his location within the questionnaire? ➢ The respondent must have a sense of being 'in control' when completing the form. ➢ Make clearly visible which entry field is active. ➢ Ensure that the respondent is able to navigate conveniently. ➢ Provide a progress indicator. ➢ Make clear which screens and questions still have to be answered.	□ □ □ □
4.	Has optimal use been made of the electronic way of presenting the questionnaire? ➢ Is the routing through the questionnaire automated? ➢ Is there a back button for people who get stuck or want to return to a previous step? ➢ Does the program issue a warning if the respondent forgets to answer a relevant question or gives improbable answers? ➢ Are all calculations automated? ➢ Does the respondent receive a confirmation of receipt after sending in the electronic questionnaire?	□ □ □ □
5.	Does the respondent receive feedback for every action? ➢ Every action must have a visible effect. Avoid surprises for the respondent.	□ □ □ □

6.	Is scrolling avoided? ➢ Horizontal scrolling in particular is extremely undesirable. ➢ If scrolling is necessary, indicate that it is possible. A scroll bar alone is not always enough.	□ □ □ □
7.	Can any necessary explanation be displayed on the screen at the same time as the question and response options? ➢ Avoid a possible pop-up with explanations obscuring the question. ➢ Avoid respondents accidentally sending in the questionnaire before it is finished.	□ □ □ □
8.	Are error messages understandable and clear? ➢ State as specifically as possible what the problem is and how it can be solved. ➢ An error message should not obscure the relevant question.	□ □ □ □
9.	Are the data saved while the questionnaire is being filled? ➢ The given responses should be saved while the questionnaire is being filled in. Mind the performance of the questionnaire programme, making back-ups takes time. Often it is not necessary to make a back-up after each response but only at specific intervals. ➢ Allow respondents to save a partly filled in questionnaire and to resume responding at a later time.	□ □ □ □
10.	For business surveys: Is it possible to print the questionnaire, before and after the questions are answered? ➢ In many cases there is more than one respondent for one business questionnaire. The contact person for the NSI will collect all the data from other colleagues on a paper version of the questionnaire before entering the data in the electronic version. ➢ Often business respondents want to have a paper print of the questionnaire for their paper archive.	□ □ □ □
11.	For business surveys: Is it possible to save an electronic copy of the questionnaire? ➢ Business respondents often not only want a paper copy, but also an electronic copy for their archives.	□ □ □ □
12.	Have the guidelines for Web Accessibility for people with impairments been taken into account? ➢ See www.w3.org/WAI/ for information on Web Accessibility Initiative – Cooperation Framework for Guidance on Advanced Technologies, Evaluation Methodologies, and Research Agenda Setting to Support eAccessibility (WAI-ACT, European Commission Project, 7th Framework Programme, IST 287725) and their guidelines.	□ □ □ □

Appendix 3. Score form for expert review of phrasing

Francine Dehue

1. Interpretation

1.1. AQ: Applicability of question: The question is not realistic enough:

1.1.1. Hypothetical or fictitious data are requested (AQH)

1.1.2. The respondent is the wrong person (AQI)

1.1.3. Future data is requested (AQF)

1.1.4. The requested data are unknown (AQU)

1.2. MC: Meaning of key concept: The key concept in the question is unclear:

1.2.1. The key concept is unknown (MCU)

1.2.2. The key concept is vague (MCV)

1.2.3. The key concept has a double meaning (MCD)

1.3. MQ: Meaning of question: The meaning of the question is unclear:

1.3.1. The meaning of the question is too unspecific (MQU)

1.3.2. The reference framework is insufficiently specified (MQR)

1.3.3. The instruction in the question is unclear or misleading (MQI)

1.3.4. The question constitutes a double question (MQD)

1.3.5. A meaning other than the one intended is attributed to the question or the response categories (MQA)

2. Task burden

2.1. LC: Logical consistency: The link between question and answer is inconsistent

2.1.1. The question contradicts earlier answers (LCR)

2.1.2. The question prompts a question-answer mismatch (LCA)

2.1.3. More than one permissible answer is possible (LCM)

2.1.4. Response categories are missing, overlap, are irrelevant or are of a different dimension from what the question suggests (LCC)

2.1.5. The question prompts interruption while reading out (LCO)

2.1.6. The phrasing of the question breaches linguistic rules (LCT)

2.2. CQ: Complexity of question: The question is too complicated

2.2.1. The question comprises too many key concepts (CQC)

2.2.2. The question comprises too many instructions (CQI)

2.2.3. The question comprises a double negative (CQN)

2.2.4. The question comprises implicit assumptions (CQA)

2.3. CL: Cognitive burden: Answering the question demands too much information processing

2.3.1. The response categories are too extensive or too difficult (CLC)

2.3.2. The answer format is too complicated (CLA)

2.3.3. It is too difficult to reach a judgment (CLJ)

2.3.4. It is too difficult to recall the desired information (CBR)

2.3.5. The answer demands complex derivations (CBD)

3. Perception

3.1. AM: Affection / motivation: The question is undesirable

3.1.1. The relevance is unclear (AMR)

3.1.2. The question is too personal (AMP)

3.1.3. The question is incomplete (AMI)

3.2. NS: Norm sensitivity: The question is not neutral and leading

3.2.1. The question comprises emotionally charged concepts (NSC)

3.2.2. The question evokes socially desirable answers (NSS)

3.2.3. The response categories are normative (NSN)

Appendix 4. Interviewer evaluation method

Wim Burgers

Objective

It is both usual and desirable to evaluate population and business surveys with interviewers. This process yields information about the survey progress, and possible or actual problems perceived by interviewers, respondents, or both. This information contributes to decisions about adjustments, structural improvements, data collection methods, and the questionnaire.

A method developed in recent years is able to yield relevant and valuable information efficiently and quickly with the cooperation of a limited number of interviewers. The output from an evaluation performed in accordance with this method is an overview of obstacles weighted according to the importance ascribed by the interviewers to the individual points.

Method

Preparation

One principle of the evaluation method is that the participating interviewers are guided as little as possible in their thinking and conclusions, and another is that all participating interviewers contribute to the evaluation. This implies that any subjects that are submitted by Statistics Netherlands employees for interviewer evaluation – unless also raised by the interviewers themselves – will not be considered until the interviewers have had their say. The list of subjects proposed by those involved in Statistics Netherlands (researcher, project leader, survey supervisor) as candidates for evaluation is prepared in advance.

Approximately 15 interviewers are invited for an evaluation. Experience has shown that, provided the group is representative of the regions and degree of urbanization, the results of an evaluation with this number of interviewers is representative for the survey staff.

Listing

The session starts by explaining to the interviewers the objective, what is to be evaluated, and how. The number of fieldwork aspects involved (e.g. contact strategy, questionnaire, survey material and logistics) depends on the survey. The group is then divided into pairs. Each pair lists the problems, complaints and unclear points they have encountered in the execution of fieldwork. They are also invited to report any conspicuously favourable matters. Each pair notes all the points mentioned in the form of a statement on a large sheet of paper. When complete, the pairs' lists of statements are put on display.

Explanation

The statements are discussed in the entire group one at a time. All attendees may ask an interviewer who wrote a statement for an explanation. It is then important to avoid discussion of the statements, which might otherwise prompt the interviewer or interviewers who wrote a statement to defend it, or its validity. In addition, every interviewer can later give a score to indicate agreement or disagreement with a statement. The objective of explaining the statements is to allow for clarification of any points, so that it is clear to all participants what the statement is intended to mean. The statement is rephrased if necessary to avoid any misunderstanding. The explanation also provides an opportunity to reduce the number of statements by combining identical or similar ones.

If necessary one of the researchers or project leaders present may provide background information as to why a given question or block of questions is in the questionnaire. Here too, this must not lead to discussion. The result is a list containing clear statements. These statements are numbered in sequence. It is established at the end of this phase whether the subjects raised in advance by Statistics Netherlands or a customer appear on the interviewers' list. Any missing statements are added to the total list. It goes without saying that these statements are explained if needed.

Scoring

Finally all interviewers are requested to give a rating to each statement, on a score form. The interviewers' scores facilitate a weighted expression of any obstacles. The interviewers scores are as follows.

The figure is between 1 and 10, where a 1 means 'disagree completely with the statement' and a 10 means 'agree completely with the statement'. Interviewers are asked in advance to be discriminating in their scoring, in the sense of not restricting themselves to ones and tens, but to think about how much they personally agree or disagree with a given statement. If they have no experience of the situation referred to in a statement, or have no opinion, they give a score of zero. (Some problems relate to a certain area e.g. major cities, which some interviewers will be unfamiliar with.)

Processing

Once scores have been given for all the statements the session with the interviewers ends. Processing of the scores can now start.

All the statements are tabulated by sequence number and then the statement scores for each interviewer entered. After entering the scores the total and average for each statement are calculated.

If the scores are to be compared, they must all point in the same direction. This sometimes necessitates a 'reversal' of the scores as part of the processing. This normally means reversing the sense of the scores of the positively phrased statements (e.g. 'This is a convenient questionnaire to work with') to be 'negative'. A score of '1' then becomes '10', '2' becomes '9', and so on.

Finally the average score is sorted in increasing and decreasing order, producing a summary of the importance attached by the interviewers to a statement.

Interviewer evaluation score form

Statement	Score		Statement	Score
1.			26.	
2.			27.	
3.			28.	
4.			29.	
5.			30.	
6.			31.	
7.			32.	
8.			33.	
9.			34.	
10.			35.	
11			36.	
12.			37.	
13.			38.	
14.			39.	
15.			40.	
16.			41.	
17.			42.	
18.			43.	
19.			44.	
20.			45.	
21.			46.	
22.			47.	
23.			48.	
24.			49.	
25.			50.	

Appendix 5a. Specimen CAPI / CATI interview observation form

Survey name	RESPONDENT										INTERVIEWER			Interviewer .... Date .... Time .... Respondent no. .... Version
	Question				Answer									
Q-no.	Doesn’t understand question	Doesn’t understand term	Too long	Lacks explanation	Answer not among available options	Doesn’t understand term	Answer categories too long	Too many answer categories	Gives multiple answer categories	Reason for refusal	Repeat question	Repeat answer	Question poorly phrased	**Comments**

Appendix 5b. Specimen CAWI questionnaire observation form

Questions Lab test Mixed Mode 2007 observation form

Interviewer	
Date	
Time	
Respondent	
Questionnaire version	

Block	Comments
Opening	
What is R's initial response to the screen?	
Does R read the introduction?	
Can R proceed to the next question?	
Other comments	
Background questions	
Can R answer the questions?	
What does R identify spontaneously as clear or unclear?	
Does R miss any explanations?	
Does R miss any response categories?	
Other comments	
Work and education	
Can R answer the questions? Can R find a category for his business and profession/job?	
What does R identify spontaneously as clear or unclear?	
Does R miss any explanations?	
Does R miss any response categories?	
How sure is R with the answers about the past?	
How sure is R with the answers about the partner?	
Other comments	

Appendix 6. Standard protocol for questionnaire evaluation company visits

Purpose of the protocol and the visit

The protocol is a guide for what should ideally be raised in a questionnaire evaluation on a company site. In practice – in particular with tests on site – some relevant aspects will not always be raised. The most important objective of an evaluation visit is to gain an understanding of the following:

- where are entry errors made?
- where does a high response burden occur?
- why do these problems arise, and what can be done about them?

This standard protocol should be augmented with specific observation points and questions for the questionnaire to be tested.

Two types of visit: observation and follow-up interview

There are two types of test visit: observation and follow-up interview. The respondent is observed while completing the questionnaire during an observation visit. Ideally the respondent works on the questionnaire as if no one from Statistics Netherlands were present – after implementation other people should complete the questionnaire in the same way. When finished, the answering process is discussed and supplementary questions may be asked. In follow-up interviews the respondent will have completed the questionnaire prior to the visit and the answering process will be reconstructed in the interview.

The advantage of an observation visit is that the place where problems arise can be established accurately. Detailed information, such as the exact location of error messages in a computer questionnaire, or the fact that a question was originally skipped, is impossible to discover after answering has ended. Observation is therefore an effective method, in particular for investigating the user-friendliness of a design. However, observation has the disadvantage that the presence of researchers can disrupt the normal answering process. Respondents may feel pressured by the observation and therefore skimp some matters, or conversely try extra hard, or resort to seeking help from the researchers instead of looking for help information on their own. Follow-up interviews therefore give a more realistic picture of the answering process, albeit less detailed. It is usually advisable to have both follow-up interviews and observations within a test.

First look and listen, then provide assistance and explanation.

For tests of business survey, unlike social surveys, test respondents will usually have experience of completing Statistics Netherlands surveys, and will complete future test and other questionnaires from Statistics Netherlands. The test visit must therefore always allow for answering the respondent's questions and correcting any entry errors found in the test. However, the main objective of the test is to explore how the questionnaire works. This means it is important that the respondent simply completes the questionnaire in the test in just the same way as he would do normally. There will be an opportunity after completing the questionnaire for any questions that arise. Researchers may provide help only if someone is really stuck, and in a normal survey would phone the help desk, for example. These basic test rules must be explained properly to the respondent in advance.

<table>
<tr><th colspan="2">Standard Protocol for test questionnaire evaluation company visit
January 2010 version</th></tr>
<tr><td>What</td><td>Objective & points for attention</td></tr>
<tr><td>1. Introduction</td><td>Introduce visitors and explain the objective and structure of the visit
▪ Objective of visit: 1. to evaluate how easy the revised questionnaire is to use 2. if applicable: to help /correct answering
▪ Structure of visit:
o general questions in advance
o complete the questionnaire (for an observation visit) OR (for a follow-up interview) map out as precisely as possible how the answering proceeded
o questions about the answering process
o any explanation of technical matters and support in answering
▪ Record permission for recording</td></tr>
<tr><td>2. General questions in advance</td><td>Dealing with Statistics Netherlands questionnaires in general and the test questionnaire in particular
▪ What kind of company is this?
▪ What position does the contact person (referred to below as CP) hold?
▪ Has the test questionnaire been completed before? (Only if it is already in the field.)
▪ Which other questionnaires does the company complete for Statistics Netherlands, and who answers them?
▪ How did the CP acquire responsibility for this questionnaire?
▪ How is it determined in the company whether and when the test questionnaire is completed?
▪ Who is responsible for answering the test questionnaire? Who is involved? If more than one person: what is the distribution of tasks?
▪ Are specific data kept for the express purpose of completing the questionnaires? If so, what data, and how much time does this involve?
▪ What does the CP think of the accompanying material (e.g. letter, e-mail, and the folder, if any)
Has CP been in contact with Statistics Netherlands about the test questionnaire? If so, about which aspect, and how did the contact proceed?</td></tr>
</table>

<table>
<tr><td>3. Observing or reconstructing the answering process</td><td>Recording how the CP completes or completed the questionnaire
1. Starting the questionnaire
How well did starting the questionnaire go? (E.g. could the CP easily log in / download / start on the correct page.)
Were any supplementary materials such as a letter, or folder consulted?
Were any introductory texts / screens read?
Did the CP browse through the questionnaire first, or start answering immediately?
2. Answering the questionnaire
For each questionnaire item observe how the CP arrives at an answer: note which items produced errors (and if possible also: what went wrong and why) and which items cause a high response burden (also indicating why).
Did the program produce warnings and error messages during the answering process? What were they? How did the CP react?
Specific points for attention for answering process. Insert test questionnaire here.
3. Sending and concluding
N.B. send computer questionnaires only if the completed questionnaire can still be opened. If not, first discuss and correct any errors in the observation visit.
How did sending proceed?
How long did it take to complete the questionnaire?
Did the CP get so severely stuck while answering that he needed help? If so: when and why did that happen?</td></tr>
<tr><td>4. CP’s opinion about the questionnaire</td><td>Ascertain CP’s opinion about the questionnaire and probe specific points for attention
What did the CP think of the content and the ease of use of this questionnaire, including both favourable and unfavourable points?
If anything in the test questionnaire needs to be changed, what would the CP prefer to be different?
If not mentioned above, what does the CP think specifically about:
the contact method (letter, folder);
downloading and installing?
navigation (index, browse buttons and OK button).
general layout (e.g. screen layout, use of colour, legibility of letters);
explanations: place and content;
menu functions;
sending?
Add to this list any specific points for attention for test questionnaires.
Ask for a printout / copy / screen dumps of the completed questionnaire (if possible also as a computer file).</td></tr>
</table>

5. Explain and correct the form	**This provides an opportunity to address content-related problems and provide help with answering** ▪ Correct items where necessary. ▪ Answer CP's earlier questions and give any explanation needed.
6. Concluding questions	**Ask about any outstanding points and close the discussion** ▪ Does the CP personally use Statistics Netherlands publications? If so, which ones? And does anyone else in the company ever use Statistics Netherlands figures? ▪ Does the CP have any questions or remarks that have not been raised so far?

Version history

Version	Date	Description	Authors	Reviewers
Dutch version: Vragenlijstontwikkeling				
1.0	24-02-2010	First Dutch version	Deirdre Giesen Vivian Meertens Rachel Vis-Visschers Dirkjan Beukenhorst	Ger Snijkers Mariëtte Vosmer
English version: Questionnaire development				
1.0E	11-10-2012	First English version	Deirdre Giesen Vivian Meertens Rachel Vis-Visschers Dirkjan Beukenhorst	