



This module is part of the

# Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

# Theme: Questionnaire Design – Main Module

## Contents

|   |    |
|---|----|
| General section.....                                      | 3  |
| 1. Summary .....  | 3  |
| 2. General description.....                               | 3  |
| 2.1 Characteristics of the establishment population ..... | 3  |
| 2.2 The response process .....                            | 4  |
| 2.3 From objectives to a draft questionnaire .....        | 5  |
| 2.4 Prototype questionnaire .....                         | 7  |
| 2.5 Systematic testing .....                              | 8  |
| 2.6 Questionnaire design standards .....                  | 9  |
| 2.7 Quality .....   | 9  |
| 2.8 Omnibus surveys .....                                 | 10 |
| 2.9 Ad hoc surveys .....                                  | 10 |
| 3. Design issues .....                                    | 10 |
| 4. Available software tools .....                         | 11 |
| 5. Decision tree of methods .....                         | 11 |
| 6. Glossary .....   | 11 |
| 7. References .....                                       | 11 |
| Interconnections with other modules.....                  | 13 |
| Administrative section.....                               | 14 |

## General section

### 1. Summary

In statistical surveys, the questionnaire is the pipeline which enables the flow of desired data. Although questionnaire design is part of the operational phase of a survey, it is critical in terms of survey objectives. It is difficult to compensate at later stages errors made due to an insufficient instrument (Brancato et al., 2006). What must be stressed is the iterative nature of its design and development. The relationship between information demand and the response burden has to be taken into account when introducing new forms and assessing existing ones. The thirst for more and more facts and figures must be balanced against the reporting unit's burden, quality aspects and costs.

As part of the survey process, the questionnaire preparation process, which is by its very nature an iterative process of improvement and development, must also be seen as a permanent and continuous cycle.

This module brings up the general issues connected with the questionnaire preparation for statistical data collection. The context is set for business surveys. The following modules are devoted to specific parts of the questionnaire design, making together with this module a list of modules devoted to the Questionnaire Design topic in this Handbook:

1. Electronic Questionnaire Design.
2. Editing During Data Collection.
3. Testing the Questionnaire.

### 2. General description

#### 2.1 *Characteristics of the establishment population*

General characteristics of establishment surveys, which distinguish them from household surveys, and further affect questionnaire design, development and testing, include:

- The response process is complex and burdensome, since preparing the required data entails a mixture of organisational and individual tasks. Data are mostly quantitative and their acquisition requires access to business records.
- The predominant role of self-administered questionnaires and the role of the respondent. In the case of establishments, the respondent is usually a representative, who acts as a data provider in an organisational environment.
- The mandatory character of reporting as opposed to mostly voluntarily aspect of household surveys.
- Concepts and definitions are technical, complex, and often based on legal or regulatory considerations. Practices, terminology and standards used by businesses in their daily operations, the accounting standards, for example, are the context that needs to be taken into account. This calls for detailed instructions that accompany questions.
- The distributions of totals of the target population are skewed in favor of larger establishments.

- Timeliness is often given priority over quality.
- The longitudinal character of surveys and overall reluctance to changes in measurement instruments used.

## 2.2 *The response process*

This section only slightly touches the question of response process models in business surveys; a more deep discussion on the subject the reader can find in the theme module “Response – Response Process”.

When a respondent is asked for information by a questionnaire, a series of activities must be performed before the task is completed. A better understanding of this task can help to make the answering process less burdensome for the respondent by applying the knowledge to improve the questionnaire.

### 2.2.1 *Traditional response process model*

The study of what happens when an interview is conducted to elicit answers to survey questions laid the foundation for the response model consisting of separate stages. The task approach, which divides the process into stages, paved the way for techniques of detecting problems with questions and improving questionnaires. The model, originating from social surveys, was developed by Tourangeau (1984) and consists of four cognitive steps: comprehension – understanding the question, retrieval – recalling the fact from memory, judgment – assessment of its correctness, reporting – formatting the response. The model provided the basis of cognitive interviewing techniques practices.

### 2.2.2 *The model for business surveys*

The starting point for the response process model in social surveys is the individual, on whom the whole model is based. Models for establishment surveys, however, must adopt a different perspective to account for the fact that the response takes place in an organisation, which constitutes a ‘universe’, with all social connections within it. Nevertheless, those models are based on the four step cognitive model. To integrate organisational and individual factors, other models have been developed, such as the Hybrid Response Process Model (Sudman, Willimack, Nichols and Mesenbourg, 2000; Willimack and Nichols, 2001, 2010), or the Multidimensional Integral Business Survey Response (MIBSR) model (Bavdaž, 2010a) which distinguishes between the business/organisational and individual/personal levels. In a study by Lorenc (2007) the Socially Distributed Cognition theory was used to study the establishment response process, whereby an establishment is treated as a unit and survey response-related processes are analysed within the framework of representational states of various interactions between persons and tools used, over time. These models can be seen as complementary ways of gaining a better understanding of processes involved in establishments survey response with a view to reducing the measurement error and obtaining valid and reliable data. What follows below is the Hybrid model, which illustrates how the original concept of cognitive steps has been developed and adapted to the needs of establishment surveys:

1. Encoding in memory/Record formation.
2. Selection and identification of respondent.
3. Assessment of priorities.

4. Comprehension of the data request.
5. Retrieval of relevant information from memory and/or existing company records.
6. Judgment of the adequacy of the response.
7. Communication of the response.
8. Release of the data.

## *2.3 From objectives to a draft questionnaire*

### *2.3.1 Objectives and concepts*

Before the initial stage of the questionnaire construction begins, objectives of the survey must be identified. Consultations with respondents regarding the information demand, translated into concepts and the resulting outline, are an essential foundation for developing an adequate measurement instrument. Concepts, such as the target population and sample design, must be determined. The response process in establishment surveys recognises the important fact that data are contained in business records maintained for business reasons. This, in turn, is related to the issue of data availability and their matching with survey concepts. Consultation studies with data users, subject data experts and survey methodologists at the early stage help to avoid the discrepancy between the intended objective and actually collected data. Exploratory studies are the way to determine the existence of data and the complexity of the process of compiling them. Most often the concepts used are complex and require technical definitions. Cooperation between parties at the interface between methodology and subject fields can lead to a better understanding of technical terms.

### *2.3.2 Variables*

Conceptual ideas must be broken down into definitions and lists of name data items, all of which leads in a straightforward way to questions. The longitudinal character of economic surveys and the goal to measure changes in time calls for the stability of variables. This strategy has two consequences. For one thing, it can ease the response burden; on the other hand, previous errors can persist in future survey periods. For this reason, changes in questionnaires should be made with caution and respondents should always be notified.

### *2.3.3 Determine data collection modes*

The data collection method chosen determines the layout of the questionnaire. This issue should be resolved before starting the design process. Two major types of data collection modes might be distinguished from the administrative point of view:

- Interviewer-assisted.
- Self-administered.

From a technological perspective they can be classified into:

- Paper-based interviewing.
- Computer-assisted interviewing.

Computer assisted interviewing, so widely used nowadays as to be considered a standard, comprises:

- Interviewer-assisted – CAPI and CATI.
- Self-administered – Web interviewing, CASI.

The data collection topic is covered in several modules on “Data Collection” in this Handbook.

### *2.3.3.1 Suitability of the interviewer-administered mode*

Personal interviewing can be considered suitable for a small sample of respondents, where concepts and questions are moderately or highly complex. Another situation where those methods can be useful are observation or panel surveys. Personal visits can be a follow-up method to mailed questionnaires. The costs of personal interviewing are the main disadvantage, as these methods are very expensive. The personal mode is more suitable for voluntary social surveys, where guidance from an interviewer is necessary to elicit more accurate responses and increase response rates.

Telephone based methods are suitable for interviews involving questions with simple concepts, where the number of items does not exceed 10. With a greater number of items, telephone based methods become problematic: questionnaires with over 40 items are regarded as unsuitable (ABS Forms manual, 2010). Similarly, the duration of the interview is an important factor when choosing a collection mode. The maximum limit for the telephone-based mode is around 20 minutes. Additional factors influencing the choice of the mode include the survey frequency and sample size. Although the front-end costs of telephone interviews are lower compared to personal visits, costs of preparing CATI instruments also keep rising, which requires a trade-off between costs and benefits. Thus, the larger the sample size and the more frequent a survey is, the more cost-effective such techniques are.

### *2.3.3.2 Suitability of the self-administered mode*

Characteristics of the response process in establishment surveys underline the role of business data records. The requested data are stored in records, which has implications for its retrieval: selecting the proper person, the knowledge of business records. Another typical challenge is the need to merge data from different departments of the institution. The use of complex terms and definitions in establishment surveys makes the response additionally burdensome. When all these aspects are considered, it becomes clear that the most suitable mode of data collection in business surveys is the self-administered mode. If this is the case, the weight of communication rests on the questionnaire and its content. Regardless of the technique applied – be that paper-based or electronic – some common elements can be distinguished in the questionnaire: it is what methodologists call languages: verbal, numerical, symbolic and graphical. Questions express the meaning of concepts, while numbers are the characteristic trait of economic surveys. Graphics and symbolic language influence the flow and cognitive burden. Generally, the visual side is the only communicative medium as far as the paper self-administered mode is concerned. In the self-administered mode the role of instructions is of the utmost importance, whenever it is necessary to clarify the meaning of complex definitions. Methodological papers advise placing instructions close to questions rather than using separate booklets. Another recommendation is to formulate instructions as questions, that is, incorporating them into questions content. Placing instructions adjacent to or within the question can improve understanding and making them available for easy reference (Tuttle et al., 2007).

### 2.3.3.3 *The mixed mode*

The burden of response in establishment surveys and the low motivation for respondents to participate in those surveys are two factors, which motivate statistical agencies to look for ways of easing the response burden and improving the response rates. The common practice is to allow respondents to choose the collection, such as mail, fax, mail out – fax in, web. A growing number of surveys supplement the main mode of collection by other methods (Nicholls, Mesenbourg, Andrews, and De Leeuw, 2000). Another example of mixed mode data collection is making an initial telephone contact to choose the proper person in the establishment as a respondent. Initial contact can also aim to confirm the identification of an establishment and to announce the upcoming survey data collection (Goldenberg et al., 1997). The actual collection is then conducted in a unimodal way and the usage of the initial mix-mode system will reduce nonresponse and has no implications for measurement errors (de Leeuw, 2005). A follow-up contact in the case of nonresponse to elicit response can be yet another example of effective multi-mode collection or sequential multi-mode.

## 2.4 *Prototype questionnaire*

Major concerns in the questionnaire design process when treated as a whole are:

- introduction – the goals of the survey, status (mandatory or voluntarily), deadline date, collection mode;
- motivation – respondent factors affect every step of the questionnaire construction: goals of the survey must be convincing to the respondent, who should also see benefits resulting from their participation;
- understanding – the logical and concise structure of questions and concepts;
- flow – layout, sections and groups should facilitate an intuitive and clear path from start to completion.

One should emphasise the iterative character of designing, developing and testing questionnaires. A typical iterative cycle starts from developing the initial measurement instrument, which is then reviewed by experts, pretested and, finally, submitted to another revision.

In an effort to work out a consistent image of the surveying agency and to improve data collection instruments that can ease the reporting burden, it is necessary to formulate guidelines for questionnaire design and development. The related aim is also to achieve a coherent “look and feel” of the data collection instrument. These guidelines distinguish several groups of elements that a questionnaire consists of (Morrison, 2007, 2008; Morrison, Dillman and Christian, 2010):

- Question wording – questions should be formulated as sentences ending with question marks, not sentence fragments to be completed; alternatively as imperatives. The question word at the beginning helps to recognise that an answer is expected. It is preferable to have a larger number of simple questions than fewer more complex ones.
- Visual design – the proposed rules can be divided into page layout guidelines and response field options. Theories about the influence of visual design on the question interpretation and comprehension suggest that some layouts contribute to questionnaires being perceived as more friendly and simpler to get through than others. For example, one column format is

easier to follow thanks to its unidirectionality, although two columns may be used for simple numerical data and paper formats. Placing too many graphics and symbols not closely related to the response task is regarded as visual clutter. One way to avoid this is the consistent use of fonts and their attributes for the same purpose throughout the questionnaire. Clutter can also result from vague organisational logic. For example, things placed close to each other seem to be related. This proximity principle can be used to indicate the beginning and the end of one question and the start of the next one. Blank spaces should be used to separate questionnaire items rather than lines, which can break up groups of elements that are, in fact, related. Groupings and separation spaces can also assist the respondent in navigating through the form. The same applies to questions and response options – questions can be linked with answer options by means of leading dots in the case of a paper questionnaire; in an electronic questionnaire shades and colours are available. It is advisable to provide a clear indication of units of required items. Any changes in the flow should be signalled by strong visual cues. Aligning questions helps to perceive the flow to be a natural consecutive path to follow, with answer options placed in one column and along a line.

- Instructions – guidelines generally suggest incorporating instructions into the questionnaire, especially in the case of business surveys, where definitions and descriptions of necessary steps to reach the value required by an item in the questionnaire are of great importance. Reference to separate documents adds burden to the response task and increases the probability of omitting an important detail. Wherever possible, instructions should be incorporated into questions. Bullet lists are more advisable than narrative paragraphs, which produce a more congested impression and require more careful reading.
- Matrices – though users in companies are accustomed to using tabulated data, such as spreadsheets, it is advisable to use tables with caution. They are burdensome and more difficult to comprehend. The decision whether or not to use a table can be made based on the pretesting phase of the questionnaire. At that stage it can be determined whether using a table will not place too much of a demand on respondents' understanding of response options. Properly designed matrices can decrease complexity: a limited number of items, clear navigational path within a table, linkage of rows and columns, using lines to indicate the direction a respondent should follow.

## 2.5 *Systematic testing*

According to principle 8 of The European Statistics Code of Practice: “In the case of statistical surveys, questionnaires are systematically tested prior to the data collection.” Systematic obligations call for standardised steps to be put into practice. Secondly, testing must be done before the collection phase. Considering all the above, the following steps could be recommended:

- Pre-field testing – differs from field methods in that special conditions are prepared to gather qualitative assessment at the early stage of the design process;
- Field testing – real environment, or rather conditions reflecting the real environment, must be met to evaluate complete questionnaires;



- Evaluation – many business surveys have a longitudinal character. This provides an opportunity for continuous assessment of questionnaires and time for improvements.

## 2.6 *Questionnaire design standards*

The systematic approach to testing defined in the Code of Practice should be linked to the overall process of questionnaire design. Standards should be applied to practices used for questionnaire design in statistical agencies. Documentation prepared for various stages of questionnaire design and the code of practices provide a coherent and clear picture of a statistical institution as seen from the respondent's perspective. Electronic data collection extends the needs for standards from the visual aspect to the testing protocol. The web data collection environment is characterised by its own dynamic. Standards create the framework for developing and programming techniques to build tools comprising specific components. Treating a questionnaire as an application composed of several components enables developers to determine standards for each component. Specifications for particular components provide standards for question types, field types, function types, validation technique types, layout types. Technical and programming tools are also subject to standardisation in terms of information technology. Guidelines for visual appearance are there to ensure a consistent "look and feel", which involves elements of the screen used for navigation, placement of additional non-content elements, rules for describing fonts, colours, size of text for questions and instructions. Standards for the respondent environment are difficult to describe due to their variability. All of this constitutes a challenge during the testing process. Web browsers are one example. It is not unusual for an application to behave differently when used with different web browsers. One way to solve this issue is to identify operating systems and browsers that most respondents use and then develop testing protocols that are compatible with those platforms.

## 2.7 *Quality*

Since quality is an essential aspect in the general framework of the "statistical process", the role of the questionnaire should be stressed in this respect. Standards established for quality reports require the inclusion of questionnaires and concise descriptions of the design and testing process. From the perspective of quality reports, three considerations can be mentioned:

- accuracy of statistical output – quality aspects concerning coverage sampling and nonresponse have drawn more attention (Willimack et al., 2004) than measurement errors. Among the sources of measurement errors are the mode of data collection and questionnaire design. The survey measurement instrument is one of the sources of measurement error, which is under direct control of the statistical agency (Bavdaž, 2010b). A good survey instrument can improve reliability and validity of statistical output.
- cost and respondent burden – the quality of respondent's estimates is inferior to a value obtained from records (Willimack et al., 1999). The required data retrieval process can be considered as one of the most burdensome elements of the response process (Willimack et al., 2010). Further, the more burdensome the retrieval process, the more inaccurate reported values are likely to be, ultimately leading to nonresponse.

- user needs and satisfaction – this requires attention at an early stage as well as during the process of assessing the statistical output: as a result, the measurement instrument is under continuous monitoring and assessment.

## 2.8 *Omnibus surveys*

The omnibus survey is a special kind of survey in which a respondent is asked questions on different topics. The goal of such a survey is to provide multi-subject information collected in a relatively short period of time and at low costs for a group of clients. To satisfy various user's needs the questionnaire must cover a number of different topics. Advantages of such a solution include:

- costs efficiency – clients are charged for questions they want to ask. The sampling and data collection costs are shared between all of them. If the user only wants to ask a few questions then doing this through an omnibus survey can be an effective way to satisfy research objectives at a reasonable cost.
- time efficiency – there is no need to organise resources for all sets of questions separately. Therefore, assuming the field schedule and frequency are flexibly planned there is a chance to get the results quicker than in a custom study.

Among the problematic issues of omnibus surveys are:

- sampling – since the sampling framework is predetermined for all clients who submitted questions for the survey it can be difficult to meet individual criteria and requirements, for example when the target population is to be “small establishments with less than ten employees”. Then the sample might not be large enough to elicit responses from the proper number of respondents for estimations.
- the multi-topic coverage – the goal of the survey is to cover several topics in one questionnaire. One topic section may affect the comprehension of the other. It is difficult to negotiate the order of questions and the impact of a sequence of questions is hard to assess. At the preparation stage different authors devise the questions separately. When the subject is changing, the user should be informed that a new topic is about to be introduced. Therefore, there is a need for an indication of subject change.
- complex questions – since a couple of subjects need to be covered each probably with a set of questions it is inappropriate to include long instructions and definitions. Complex questions introducing skip patterns and many multiple choices are not suitable.

## 2.9 *Ad hoc surveys*

An ad hoc survey is a survey without any plan for repetition. It is also possible to add ad hoc questions to a questionnaire used in a regular survey. Ad hoc modules included in questionnaires play their role as complements to the main modules. Incorporating additional modules creates an opportunity to provide data on different subject or specific parts of the survey subject. On the other hand adding other modules increases burden imposed on respondents, which may affect the quality of responses.

## 3. **Design issues**

#### **4. Available software tools**

#### **5. Decision tree of methods**

#### **6. Glossary**

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

#### **7. References**

ABS (2010), *Forms Design Standards Manual*. Australian Bureau of Statistics.

Brancato, G., Macchia, S., Murgia, M., Signore, M., Simeoni, G., - Italian National Institute of Statistics, ISTAT, Blanke, K., Körner, T., Nimmergut, A., - Federal Statistical Office Germany, FSO, Lima, P., Paulino, R., - National Statistical Institute of Portugal, INE, and Hoffmeyer-Zlotnik, J. H. P., - German Center for Survey Research and Methodology, ZUMA (2006), *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System*.

Bavdaž, M. (2010a), The multidimensional integral business survey response model. *Survey Methodology* **36**, 81–93.

Bavdaž, M. (2010b), Sources of measurement errors in business surveys. *Journal of Official Statistics* **26**, 25–42.

Goldenberg, K., Levin, K., Hagerty, T., Shen, T., and Cantor, D. (1997), Procedures for reducing measurement error in establishment surveys. Presented at the American Association for Public Opinion Research, Norfolk, Virginia, May 1997.

de Leeuw, E. D. (2005), To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics* **21**, 233–255.

Lorenc, B. (2007), Using the Theory of Socially Distributed Cognition to Study the Establishment Survey Response Process. Paper presented at the ICES-III, June 18-21, 2007, Montreal, Quebec, Canada.

Morrison, R. (2007), Towards the Development of Establishment Survey Questionnaire Design Guidelines at the U.S. Census Bureau. Paper presented at the ICES-III, June 18-21, 2007, Montreal, Quebec, Canada.

Morrison, R. (2008), Writing Revising Questionnaire Design Guidelines. Component of Statistics Canada Catalogue no. 11-522-X Statistics Canada’s International Symposium Series: Proceedings.

Morrison, R., Dillman, D., and Christian, L. (2010), Questionnaire Design Guidelines for Establishment Surveys. *Journal of Official Statistics* **26**, 43–85.

- Nicholls II, W. L., Mesenbourg Jr., T. L., Andrews, S. H., and De Leeuw, E. (2000), Use of New Data Collection Methods in Establishment Surveys. *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association, Buffalo.
- Sudman, S., Willimack, D. K., Nichols, E., and Mesenbourg, T.L. (2000), Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies. *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association, 327–337.
- Tuttle, A. D., Morrison, R. L., and Willimack, D. K. (2007), From Start to Pilot: A Multi-Method Approach to the Comprehensive Redesign of an Economic Survey Questionnaire. Presented at the Federal Committee on Statistical Methodology Research Conference, Arlington, VA.
- Tourangeau, R. (1984), Cognitive science and survey methods: a cognitive perspective. In: Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, National Academy Press, Washington, DC.
- Willimack, D. K., Nichols, E., and Sudman, S. (1999), Understanding the questionnaire in business surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 889–894.
- Willimack, D. and Nichols, E., (2001). Building an Alternative Response Process model for Business Survey, *Proceedings of the Annual Meeting of the American Statistical Association, August 5-9*.
- Willimack, D. K., Lyberg, L., Martin, J., Japac, L., and Whitridge, P. (2004), Evolution and Adaptation of Questionnaire Development, Evaluation, and Testing Methods for Establishment Surveys. In: Presser, S., et al. (eds.), *Methods for Testing and Evaluating Survey Questionnaires*, Chapter 19, Wiley, New York.
- Willimack, D. and Nichols, E. (2010), Hybrid Response Process Model for Business Surveys. *Journal of Official Statistics* **26**, 3–24.

## **Interconnections with other modules**

### **8. Related themes described in other modules**

1. Questionnaire Design – Electronic Questionnaire Design
2. Questionnaire Design – Editing During Data Collection
3. Questionnaire Design – Testing the Questionnaire
4. Data Collection – Main Module
5. Response – Response Process
6. Response – Response Burden

### **9. Methods explicitly referred to in this module**

- 1.

### **10. Mathematical techniques explicitly referred to in this module**

- 1.

### **11. GSBPM phases explicitly referred to in this module**

1. Sub-process 3.1

### **12. Tools explicitly referred to in this module**

- 1.

### **13. Process steps explicitly referred to in this module**

- 1.

## Administrative section

### 14. Module code

Questionnaire Design-T-Main Module

### 15. Version history

| Version | Date       | Description of changes                    | Author        | Institute  |
|---------|------------|---|---------------|------------|
| 0.1     | 04-03-2012 | first version                             | Paweł Lańduch | CSO Poland |
| 0.2     | 09-04-2013 | second version                            | Paweł Lańduch | CSO Poland |
| 0.3     | 18-09-2013 | third version                             | Paweł Lańduch | CSO Poland |
| 0.3.1   | 18-12-2013 | minor improvements                        | Paweł Lańduch | CSO Poland |
| 0.4     | 21-01-2014 | improvements after EB review              | Paweł Lańduch | CSO Poland |
| 0.4.1   | 22-01-2014 | preliminary release                       |               |            |
| 1.0     | 26-03-2014 | final version within the Memobust project |               |            |
|         |            |   |               |            |

### 16. Template version and print date

|                       |                         |
|-----------------------|-------------------------|
| Template version used | 1.0 p 4 d.d. 22-11-2012 |
| Print date            | 21-3-2014 17:26         |