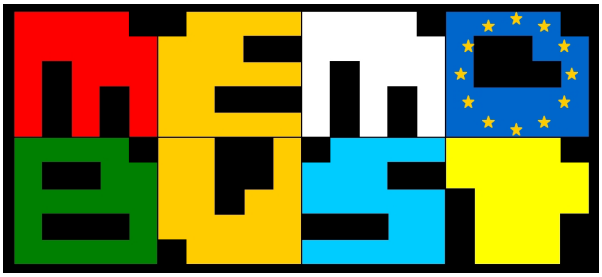


This module is part of the

# Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

# Theme: Testing the Questionnaire

## Contents

# General section

## 1. Summary

Establishment surveys differ from household surveys. This fact is reflected in the different culture of questionnaire development, evaluation and testing. First of all, the response process is more complex than in household surveys. The extensive adaptation of cognitive methods in social questionnaire development and testing based on Tourangeau's (1984) response process model, in the case of establishments, had to be enhanced by including new dimensions. The four step model, consisting of comprehension, retrieval, judgment and communication focuses on the individual. In the context of establishment surveys, however, the respondent is an informant selected within an organisation. Besides cognition, testing must also take into account the institutional frame, the need for cooperation, and the fact that required data items may be stored in business records. The non-existence of data in business records has to be taken into account as well.

Another consideration in the field of questionnaire testing is the specific nature of business surveys with their technical and intricate concepts and definitions; therefore, comprehension depends considerably on instructions. Closely connected with that and having its own consequences is the predominately self-administered data collection mode. There are other distinguishing features of the establishment population which pose a challenge for testing procedures. Among those are the longitudinal character of business surveys and the subsequent use of resulting data as inputs for other surveys. Yet another problem is a negative attitude to any changes in the questionnaires.

Nonetheless, the need for testing is beyond contention. This requirement is stated clearly in the Eurostat Code of Practice: "In the case of statistical surveys, questionnaires are systematically tested prior to the data collection". Ongoing data collection instruments are also under scrutiny. The goals to achieve include the improvement of the quality of statistical output, the reduction of costs to the surveying agency and to respondents, a decrease in the scope of output variables, an increase in the use of administrative data (Giesen, 2005). The adoption of Computer Assisted Interviewing has prompted redesign efforts to explore new prospects in data collection. This has added a new level of complexity to questionnaire testing. One new dimension is usability testing, which is intended to assess if the testing tool is user friendly and whether the interaction with the computer is intuitive and simple for the respondent. The optimal approach to efficient testing requires the involvement of end users. This leads to a paradox: in an effort to improve the collection instrument and ease the response burden, another burden is imposed on respondents (Willimack, 2005). Adding more burden during the response process, which in itself is burdensome, can hardly meet with the respondent's approval. It is, therefore, important that respondents should be aware of the goals of the testing procedure, which is intended to simplify and ease the response. When the aims of the procedure are clear, additional efforts can be received with a higher degree of approval. On the other hand, the iterative and longitudinal character of business surveys makes it possible to work out a systematic approach to improving the data collection instrument in a step-by-step procedure, which involves incorporating the testing and developing research into ongoing and repeated surveys. Instead of the usual practice of relying on post-collection activities to correct errors, a new paradigm is proposed, encouraging research on improving questionnaire design that leads to "error prevention rather than error correction" (Willimack et al., 2004).

A broad spectrum of recommended practices for developing and testing statistic questionnaires can be found in the Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System (2006). The Handbook adds valuable enhancement to the general subject of developing and testing questionnaires in statistical surveys. A detailed discussion of testing and evaluation questionnaires for establishment surveys can be found in Willimack (2013).

## 2. General description

### *2.1 Iterative itinerary*

The development, testing and evaluation of questionnaires is part of a continuous process consisting of separate but linked stages along a continuum (Goldenberg et al., 2002). The process can be divided into 2 parts. The first part, including development and testing, and the second part, which comprises the assessment of measurement instruments after they have been used in the data collection process. The path goes through iterative steps, often going back and repeating the same cycle again. The starting point is to work out the survey goals. A draft questionnaire is a translation of concepts and definitions into questions and variables. Considering the precise and complex nature of concepts in establishment surveys the role of subject data experts cannot be neglected. The requirements and objectives of the survey should be combined with design aspects such as layout, technical constrains, instructions. A systematic approach requires guidelines in order to guarantee a consistent "look and feel" of the questionnaire. Draft versions have to be submitted to questionnaire design experts and subject data experts for reviewing. Data users should also play an active role in reviewing questionnaires because of the technical nature of the economic data and their stringency (Willimack et al., 2004). This stage is followed by the pretesting phase. Newly developed questionnaires require several rounds of pretesting. The pretesting process is limited by the costs and time. Beginning with internal staff testing, efficient testing should also involve end users. Findings from pretesting are the basis for further revisions. Ongoing surveys questionnaires can also be submitted for evaluation. The questionnaire may require a redesign to reduce costs, response burden, or change the data collection method. The testing and evaluation process should be based on the establishment response process model (Snijkers et al., 2005). The whole designing and redesigning process is iterative and open.

### *2.2 The response process model as a tool for testing the questionnaire*

The module "Response – Response Process" provides a general discussion about response process models in business surveys. In this place the response process is referred only as the foundation for the cognitive approach to improve and evaluate questionnaires in statistical measurement. The knowledge of consecutive phases the respondent may go through starting from the decision to participate in a statistical survey to its successful completion can reduce the burden and, consequently diminish the measurement error. The response process in establishment surveys is more complex than in social surveys, which makes the matter even more important. Therefore, the response process can be a tool to evaluate the questionnaire and to find ways to improve it. An example of such an implementation can be found in a paper by Giesen (2007).

### *2.3 Pretesting*

Pretesting involves applying testing techniques before the measuring instrument is used in the field in a survey operational stage.

### *2.3.1 Cognitive pretesting*

Cognitive aspects of survey methodology (CASM) provided the basis for extending the interviewing process, typically understood as a way of eliciting answers to questionnaire questions, to a new field of exploration, namely to survey questionnaire testing, in an attempt to reduce measurement error. Cognitive interviewing focuses on the individual who becomes "the subject" of research and on the response thought process which drives the response (Willis, 1999). The general response process model by Tourangeau (1984), who developed its conceptual theoretical framework, consists of four steps: comprehension – associated with question understanding, retrieval – recalling the required information from memory, judgment – decision about the adequacy of the response, reporting – mapping the response to question categories. Cognitive interviewing relies on two methods: think aloud, where the subject is instructed to verbalise the process of arriving at an answer, and verbal probing, where, after having answered the question the subject is "probed" to get to the bottom of the response process. These methods were adapted to social surveys where the respondent means an individual. Cognitive pretesting in establishments surveys derives from the same methods but contains significant differences. First, the response process is much more complex. The four cognitive steps are extended by including additional steps unique to the establishment culture (Sudman et al., 2000). The next important aspect is that required data are contained in business managerial systems, that is in business records, not in a person's memory. Such data can be dispersed in various parts of a company and gathering them requires cooperation between many persons. The largely quantitative nature of data is another characteristic feature. Building a protocol for cognitive pretesting should take into account all those aspects. Cognitive interviews take place at business locations rather than in laboratory conditions (Goldenberg et al., 2002; Willimack, 2008) owing to respondents' unwillingness to participate in interviews outside their workplace. What matters here as well is access to records. The process of filling the questionnaire by the respondent during a cognitive interview is complemented by assessing how business records match the required data items. The interview also focuses on the timeline of data requirements and data availability (Goldenberg et al., 2002). The complex nature of the response in business surveys requires the expansion of cognitive interviewing (Freedman and Rutchik, 2002), which consists of pre-survey design visits and cognitive testing of the questionnaire at a business location. Pre-survey design visits should test data availability, record keeping practices, the compatibility between the time data are available and the time they are submitted to a statistical agency, the need for data confidentiality. A data model and a draft questionnaire are then cognitively pretested. "Think aloud" interview cognitively tests questions, instructions and concepts used. The informal unstructured part is used to discuss business records and the questionnaire itself.

### *2.3.2 Studies of business records*

The first step in Tourangeau's response process model (1984), that is *encoding in memory*, has been complemented in the business response model by *record formation* step (Edwards and Cantor, 1991; Sudman et al., 2000). The step stresses the fact that required data are contained in business systems. This affects the further steps of the response process, such as the selection of the proper respondent with access to data records and the knowledge of those records. Survey questions might be comprehensible but the required data may not be available in business records (Willimack, 2008). Record studies in companies, conducted as pre-survey design visits, can be a useful tool to collect information about the availability of requiring data, the compatibility of record keeping practices with

data collection instruments or the burden connected with retrieving those data from managerial systems (Murlow et al., 2007). Interviewing has a cognitive background. Interviews conducted in the four subsequent stages evolve from initially being focused on the overall goals of the survey, such as concepts and definitions or organisational aspects, to aspects directly connected with records. Valuable findings from such studies provide the knowledge about many aspects of record keeping in establishments. The study by Murlow et al., showed that different data are kept in different places of an establishment and that different people have different degrees of access to company data. Since the confidentiality of company data is a crucial aspect, it is easier to get general information without looking into details. The awareness of many aspects of the survey, Research and Development Survey as a result of the study by Murlow et al., helps to rebuild the questionnaire structure. The results proved to be worth the costs and efforts.

### *2.3.3 Usability testing*

The usability theory stresses cognitive emotional aspects of the communication process and establishes principles of how to make things easier to use. While in the case of the paper questionnaire only visual improvements are possible, electronic questionnaires can be tested not only to assess their visual aspect, but also the user interface – the interconnection between the user and the computer program – and its functionality.

The theoretical background and empirical studies provide guidelines for visual elements of the questionnaire, such as consistence in the use of colours, fonts, spaces for questions and answers, answer options. Draft screens or their paper specimens submitted for assessment can develop design standards. Usability principles and the design of questionnaires are combined in the form of heuristic reviewing principles. A suite of usability tests addresses such aspects as navigation, skip routes, data entry or error correction. A product is usually subject to internal testing, before undergoing site testing by actual respondents. On-site testing, such as observation, can be connected with other cognitive testing methods. The Internet collection mode enables the application of software that records the response process in a real environment. System data logs can store information about user practices and the amount of time spent on the work with the questionnaire. This can provide additional knowledge about the response process.

## *2.4 Field testing*

Field testing differs from pretesting in that it is applied to data collection instruments after they have been used in an operational stage (i.e., when the data have been collected). However, the term is sometimes used interchangeably with pretesting. Field testing can take the form of pilot tests that can be run before the data collection phase or after data collection is complete. What differentiates them from pretests is the greater number of respondents taking part in tests; as a result, the sample scope permits statistical inference; another difference is the iterative character of pretests as opposed to a one-off administration of a pilot test (Willimack et al., 2004). Post-collection questionnaire assessment is also referred to as questionnaire evaluation (QE).

### *2.4.1 Pilot tests*

For a new or a redesigned survey questionnaire, a formal pilot test is the final step seen as a “dress rehearsal” of the measurement instrument (Goldenberg et al., 2002). During this test, all the steps of

the data collection process can be assessed. Leaving the respondent alone with a self-administered questionnaire, without the presence of an interviewer, mirrors the real environment of the response process. Evaluation of a redesigned instrument can be conducted by addressing the pilot questionnaire to a subsample of the target population and using the old form with the rest of the sample (Tuttle et al., 2010). In this particular study the results were gathered by debriefing respondents completing the pilot form and from additional questions designed to assess respondents' attitudes to the new questionnaire and by comparing it with the old one. A pilot questionnaire embedded in the final collection instrument is an opportunity to evaluate the proposed changes to the questionnaire (Willimack, 2008). This helps to avoid obstacles posed by traditional methods of improving questionnaires. For example, an additional question can be included to obtain an evaluation of how labour-intensive a questionnaire is (i.e., how much time is required to complete it).

### *2.4.2 Debriefing respondents*

After the required data have been collected, the respondent is contacted one more time. The goal of such a contact is to acquire an assessment of the quality of the gathered data or to evaluate the questionnaire itself. Findings can help to improve the data collection instrument. There are formal and informal methods of conducting debriefings. Formal methods include *response analysis surveys* (RAS). Evaluations are conducted using structured questionnaires which contain questions about data sources and response strategies (Willimack et al., 2004). The feature that differentiates them from pretests is that contacted respondents are chosen from among original survey respondents and debriefings are done after data collection has been completed. This enables generalisations and in the case of ongoing surveys, allows future revisions (Goldenberg et al., 2002). The renewed contact can be performed in person, by telephone or by means of an evaluative questionnaire. Debriefings can also have an informal or ad hoc character. The purpose of a recontact is to get feedback from respondents on questions and questionnaire elements. Findings from a small sample of respondents can provide suggestions as to which elements of the questionnaire should be changed to address reported problems (Willimack et al., 2004). General issues concerning the response process can be addressed using unstructured interviews.

### *2.4.3 Debriefing survey staff*

Survey operational personnel can be a valuable resource of knowledge about question-related problems. As intermediaries between respondents and the survey agency, offering help with questionnaire completion and recontacting respondents to resolve data item failures, they have the necessary knowledge to evaluate questionnaires and suggest potential revisions. Debriefing can be conducted both using formal methods, such as focus groups, or through informal methods (Goldenberg et al., 2002). One example of an informal method is observation of conversations between survey technical staff and respondents. Cognitive methods used in social surveys can be applied (Willimack et al., 2004). Data collection instruments can also be evaluated by data checking staff, who observe problems with data editing and who have direct contact with respondents. A cognitive debriefing session is a qualitative method of investigating problems with variables and reasons for problems (Hartwig, 2009). It involves an interview based upon a structured protocol which addresses the overall aspects of the survey and the questionnaire and its elements to improve the questionnaire or the design process. Unstructured discussions about experiences from field work can

also take place. At minimum costs, staff debriefings help to identify many problems and are the basis of long–term work on improving data collection instruments.

### *2.4.4 Post-collection data evaluation*

The need for data comparability in long time series and their subsequent usefulness in constructing economic indicators limits the scope of changes that can be introduced in questionnaires and explains the general reluctance to any changes. Post collection data analysis is routinely conducted to assess data quality (Willimack et al., 2004). Measuring non-response to questionnaire data items, detecting outliers in collected data, the rate of imputation and examining data editing failures are various methods used to evaluate how the questionnaire works in the field (Goldenberg et al., 2002). A large number in non-response items may indicate problems with data availability. Data collected by a questionnaire in a survey can be compared with data from other sources to assess data consistency and quality. Data collection analysis is also useful in assessing whether changes made in questionnaires have improved the quality of collected data (Willimack, 2008). Questionnaire pretesting can lead to changes in questionnaires. Cognitive qualitative pretesting tries to improve questionnaire understanding and ease the burden imposed on respondents. By applying quantitative methods to data collected before and after the changes it is possible to find objective measures of how these changes have worked and whether there is any improvement in data quality. The non-response rate and data edit failures can be example of such measures.

## *2.5 Standards*

Thorough questionnaire testing is a complex and burdensome process. The complex response process in establishment surveys makes the procedure even more challenging. Efficient testing requires the involvement of end users. By setting standards for this process one establishes goals that questionnaires must meet before they are actually used in data collection. DeMaio (2005) provides an example of levels of criteria a data collection instrument is expected to fulfil. The criteria are established for testing questionnaires as well as for other survey-related materials, such as introductory letters and supplemental instructions. The recommendations describe questionnaire requirements both for new surveys and redesigned measurement instruments in social and economic surveys and censuses. The minimal level is testing for proper administration of the questionnaire by an interviewer or by an end user and whether the questions are understandable. Additional recommendation extends requirements for data of great importance, and supplemental materials that are related to the survey. Self-administered electronic questionnaires further require testing that all the components of the software system behave properly and according to the design. Developing and applying guidelines and best practices for layout, visual design, field types, error handling can facilitate the testing process.

## *2.6 Software*

Testing is the process of analysing software to detect differences between the expected and actual state, and to evaluate individual pieces of software. The content of this section was mainly founded on the Certified Tester Foundation Level Syllabus, which is aimed at anyone involved in software testing (ISTQB, 2010).

The main objective of testing is to find defects and errors in a questionnaire, software or documentation. Rigorous testing of software and documentation can reduce the risk of failure in

a production environment and contribute to a high quality product. It is necessary that any defects found in the process should be corrected before allowing the system to operate in a production environment.

- **Testing reveals errors**

  Testing may indicate that there are defects, but we are not able to prove through testing that there are no errors. Testing reduces the probability that defects remain unidentified in the software, but even if no defects are found, it is no proof of software correctness.

- **Thorough testing is impossible**

  Testing everything (all combinations of inputs and preconditions) is possible only for very trivial cases. By defining the scope of tests, instead of focusing on thorough testing, we focus on risks and priorities.

- **Early testing**

  Testing procedures should start as early as possible in the software development cycle. They should also be geared to achieving the defined objectives.

- **Accumulation of errors**

  Most defects found during testing prior to release or software failures revealed during production are located in a small number of modules.

- **Pesticide paradox**

  If the same tests are repeated continuously on the same set of test cases, no new errors are found. This phenomenon involves the development of resistance to software testing (pesticide paradox[1]). To overcome this paradox, test cases must be regularly reviewed and revised. In order to check other parts of software or system to potentially find more errors, one should use new tests.

- **Testing is context-sensitive**

  Testing is done differently in different situations. For example, systems critical for safety are tested differently than e-commerce systems.

- **A false notion of correctness**

  Finding and removing errors does not help if the system is not suitable for use and does not meet users' needs and expectations.

### *2.6.1 The testing process*

The most visible part of testing is conducting tests. However, for tests to be efficient and effective, test planning should also take into account the time spent on test planning, designing test cases, preparing to perform tests and evaluating test execution status.

The basic test process consists of the following main steps:

[1] An analogous phenomenon – insects become resistant if one keeps applying the same insecticide.

– **Planning and supervision**

Test scheduling verifies the test mission, defines the objectives for testing and methods of achieving them. Test supervisions involves repeated comparison of actual testing progress and reporting with the plan and providing information about any deviations.

– **Analysis and design**

Test analysis and design are aimed at transforming general testing objectives into tangible test conditions and test designs.

– **Implementation and execution**

Test implementation and execution are a stage in which test conditions are transformed into test cases and test environment is created.

– **Assessment of the degree of completion and reporting**

At this stage, tests are evaluated in terms of predefined goals and exit criteria and the results are reported. It is specified whether more testing is needed or a change of termination conditions is necessary. Also the final report of the testing process is created.

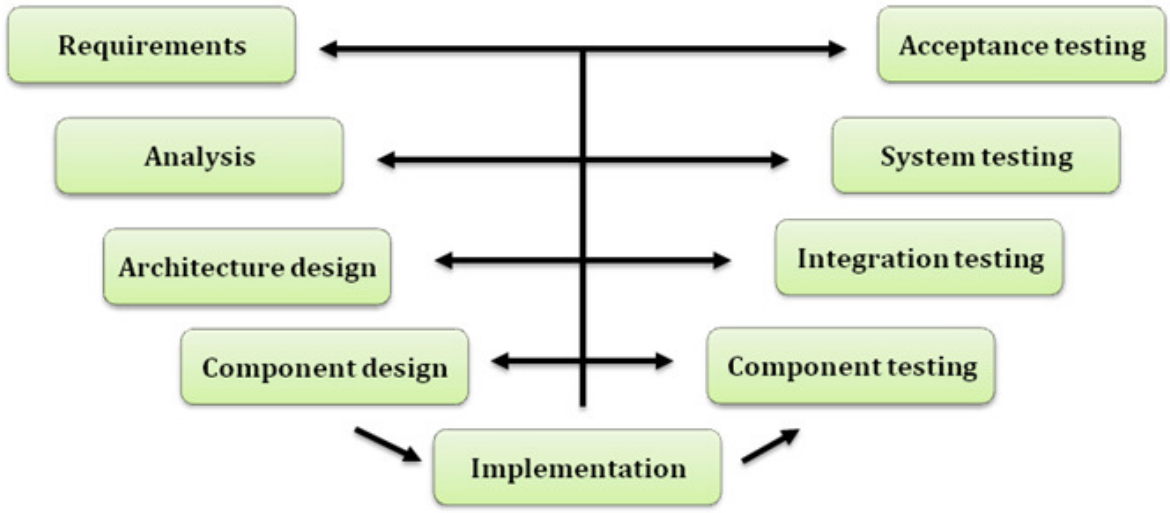
– **Test closure**

As part of closing the testing process, data are collected from completed test activities for future reference.

Although these activities are logically sequential, the process may overlap or occur simultaneously.

### *2.6.2 The testing phase in software life cycle.*

Testing does not make sense in isolation from software development activities they are connected with. Typically, four levels of tests are identified, which correspond to four levels of software:




– **Component testing** (called unit test, module test) is a programming method for testing software by performing tests that verify the functionality of individual components (units) of the program - for example, methods or objects in object-oriented programming and procedures in procedural programming.

- **Integration testing** is performed to detect errors in the interfaces and interactions between modules (assembly testing). For example, we test communication between the module that stores and provides a set of parameters and a module that uses these parameters for initiation, for example, to fill form fields with default values.
- **System testing** is intended to determine whether an integrated system already meets the functional requirements and system requirements contained in the specification.
- **Acceptance testing** is not aimed at detecting errors but obtaining a formal confirmation of software quality.
- **Regression testing** is performed to ensure that the application works after modifications, error correction or expansion (new features). This kind of testing, due to its repeatability, lends itself to automation and can reveal previously undiscovered bugs.

#### *2.6.2.1 Functional tests (black box)*

Functional tests are based on functionalities and can be made at each level of testing. It is assumed that the tester doesn't know the structure of the program or its code. Their main features are:

- no prior knowledge of the application structure is required
- data are divided into classes of equivalence
- their purpose is to test the final functionalities

#### *2.6.2.2 Structural tests (white box)*

Structural testing can be performed at all levels of the test, but its main use is to test modules and module integration. Their task is to test those parts of the design which have not been tested by functional tests. They are based on the architecture of the application. Their main features are:

- knowledge of the application structure is required
- they cannot be used to reveal missing functionalities
- their purpose is to test the application structure

#### *2.6.2.3 Non-functional tests*

Non-functional tests include performance tests, load tests, stress tests, usability tests, cooperation tests, service tests, reliability tests and tests of the ability to work across different platforms. Tests of this type determine how the system works. Their main features are:

- they assume knowledge of the application configuration
- they require multiple test platforms
- they check performance

### *2.6.3 Automated software tests*

#### *2.6.3.1 What do automated tests deliver?*

Automated tests are tests carried out with the help of specialised software. They are used to speed up the testing process, allowing you to generate test data and expected results, perform a set of tests with a final evaluation being positive or negative.

The advantages of automatic testing:

- efficient verification of bug fixes
- reuse of prepared tests
- quick reports
- comprehensive analysis of test results
- the use of large volumes of test data
- reduction in the cost of testing

Automatic testing can detect errors in the early stages of software development and protects against re-creation of the same error, which reduces the cost of creating questionnaires. Therefore, automated tests are performed at every stage of the project and invest in software testing.

Many tools are available on the market that make test preparation easier and faster. Presented below are the most popular ones.

#### *2.6.3.2 Tools used for automated testing*

Software testing varies depending on types of tests performed:

##### *2.6.3.2.1 Functional testing*

Functional tests are designed to test specific functionality including testing of the user interface.

They are used to simulate user behaviour and test questionnaire responses to these behaviours. These tests can be extremely useful at the early stages of the project. Automated tests of this type can be used as a key component of regression testing, especially in complex questionnaires, where manual testing of all functionalities is very time consuming and thus expensive. Functional tests must specify global standard methods of performing tasks such as filling out forms, login procedures etc. for future use in a single line of code that is not duplicated. Thus, for example, a change in the login procedure only requires a revision of a few lines in the log handler rather than making the same changes repeatedly at each test that required login.

Functional tests can be created either by authors of the questionnaire or target users, allowing quick and inexpensive testing under realistic conditions. Tools to enable this type of testing in the case of web questionnaires include **Selenium, Neoload** and **Rational Functional Tester**.

##### *2.6.3.2.2 Unit testing (structural)*

Unit tests are used to test individual modules of the questionnaire.

A program that performs unit tests verifies the accuracy of data input and output of the questionnaire software and the accuracy of the data processing method. It also checks integration between the modules themselves. It does not allow testing graphic elements of the questionnaire, it only serves to test the logical layer. Since unit tests focus on the logic of the questionnaire, they can only be created by programmers with high technical expertise and knowledge of the code. An example of a tool used to perform unit testing is **NUnit** for NET framework and its counterpart **JUnit** for Java and other tools of the **XUnit** family. In NUnit different classes or groups of classes are tested.

*2.6.3.2.3 Non-functional testing*

Non-functional testing includes, among others, load, performance, usability testing, and operations of the questionnaire on different platforms. These tests determine the hardware requirements and the help desk support necessary to prepare a questionnaire for action in difficult conditions, such as a very large number of users. They are expensive tests because they require experts in various fields (developers, service engineers, testers, users target). Tools to carry out non-functional tests include **NeoLoad, LoadRunner, HP LoadRunner**.

## 3. Design issues

In the context of questionnaire design one needs to mention systematic errors. While business surveys usually operate on a smaller number of variables than social surveys, their definitions tend to be complex and technical. Information in businesses data systems is organised to help companies achieve business goals but also to meet regulatory requirements. Therefore, their own definitions sometimes do not match those used by statisticians. In order to diminish problems resulting from these differences, we selectively mention two of the many design issues, when developing and testing questionnaires in business surveys:

- Using top-down approach together with bottom-up approach – the theory driven approach, with questions based on theoretical constructs, should be accompanied by explorations of data using by businesses;
- "Borrowing" questions from other surveys – in the light of the questionnaire testing, adapting a question from another survey may save the resources for additional testing, since the testing procedure has been already performed. The feasibility of "borrowing" depends on many factors of the survey design but determining it can be the first step in constructing a new survey or redesigning the current one.

## 4. Available software tools

Section 2.6.3.2 provides a couple of available software tools for various types of automatic software testing.

## 5. Decision tree of methods

Building the decision tree of methods can start with reviewing the contingency table of methods and steps in questionnaire developing, testing and evaluation. A paper by Goldenberg et al. (2002) gives an example of such a table where rows contain various methods used by statistical organisations and columns correspond to the steps specified by the method for questionnaire development, testing and evaluation (QDET).

## 6. Glossary

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

## 7. References


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, 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*.

Edwards, W. S. and Cantor, D. (1991), Toward a Response Model in Establishment Surveys. In: P. P. Biemer et al. (eds.), *Measurement Error in Surveys*, John Wiley & Sons, New York, 211–233.

Freedman, S. and Rutchik, R. (2002), Establishments as Respondents: Is Conventional Cognitive Interviewing Enough? *Presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods, Charleston, SC*.

Giesen, D. (2005), Results of the Evaluation and Redesign of the Dutch Structural Business Statistics Questionnaires. *Proceedings of the 5th QUEST Workshop, 19-21 April 2005*, Statistics Netherlands, Heerlen.

Giesen, D. (2007), The Response Process Model as a Tool for Evaluating Business Surveys. *Proceedings of the Third International Conference on Establishment Surveys (ICES-3), 18-21 June, Montreal, Canada*, American Statistical Association, Alexandria, VA, 871–880.

Goldenberg, K. L., Anderson, A. E., Willimack, D. K., Freedman, S. R., Rutchik, R. H., and Moy, L. M. (2002), Experiences Implementing Establishment Survey Questionnaire Development and Testing at Selected U.S. Government Agencies. *Presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods, Charleston, S.C., November, 2002*.

Hartwig, P. (2009), How to use edit staff debriefings in questionnaire design. *Presented at the European Establishment Statistics Workshop - EESW09*.

ISTQB (2010) International Software Testing Qualifications Board (ISTQB®) Certified Tester Foundation Level Syllabus, Version 2010 by Thomas Müller (chair), Armin Beer, Martin Klonk, Rahul Verma.

DeMaio, T. (2005), Standards for Pretesting Questionnaires and Survey Related Materials for U.S. Census Bureau Surveys and Censuses.

Murlow, J. M., Freedman, S., and Rutchik, R. (2007), Record Keeping Studies – Love 'Em or Leave 'Em. *Paper presented at the Third International Conference on Establishment Surveys - ICES-III, June 18-21, 2007, Montreal, Quebec, Canada*.

Snijkers, G., Onat, E., and Tonglet, J. (2005), The Dutch Annual Business Inquiry: Developing and testing the electronic form. *Proceedings of the 5th QUEST Workshop, 19-21 April 2005*, Statistics Netherlands, Heerlen.

Sudman, S., Willimack, D. K., Nichols, E., and Mesenbourg Jr., T. L., (2000), Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies. *Proceedings of*

*the 2nd International Conference on Establishment Surveys*, American Statistical Association, Alexandria, VA, 327–335.

Tourangeau, R. (1984), Cognitive science and survey methods: a cognitive perspective. In: T. Jabine, M. Straf, J. Tanur, and R. Tourangeau (eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. National Academy Press, Washington, DC.

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*.

Willimack, D. K. (2005), The Paradox of Respondent Burden in Testing Electronic Instruments for Establishment Surveys. *Proceedings of the 5th QUEST Workshop, 19-21 April 2005*, Statistics Netherlands, Heerlen.

Willimack, D. K. (2008), Issues in the Design and Testing of Business Survey Questionnaires: What We Know Now that We Didn't Know Then – and What We Still Don't Know. *Proceedings of the Conference on Reshaping Official Statistics*, International Association on Official Statistics, Shanghai, China.

Willimack, D. K., Lyberg, L., Martin, J., Japec, L., and Whitridge, P. (2004), Evolution and Adaptation of Questionnaire Development, Evaluation, and Testing Methods for Establishment Surveys. In: S. Presser et al. (eds.), *Methods for Testing and Evaluating Survey Questionnaires*, Wiley, New York, Chapter 19.

Willimack, D. (2013), Methods for the Development, Testing, and Evaluation of Data Collection Instruments. In: G. Snijkers, G. Haraldsen, J. Jones, and D. Willimack (eds.), *Designing and Conducting Business Surveys*, Wiley, Hoboken, New Jersey, 253–301.

Willis, G. (1999), Cognitive interviewing: a "how to" guide. Research Triangle Institute, http://appliedresearch.cancer.gov/areas/cognitive/interview.pdf.

## Interconnections with other modules

### 8. Related themes described in other modules

1. Response – Response Process

### 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.

### 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-Testing the Questionnaire

## 15. Version history

| Version | Date | Description of changes | Author | Institute |
|---|---|---|---|---|
| 0.1 | 23-02-2012 | first version | Magdalena Homenko | GUS (Poland) |
| 0.2 | 29-05-2013 | major revisions (reviews from Netherlands and Sweden) | Magdalena Homenko<br>Paweł Lańduch | GUS (Poland) |
| 0.3 | 30-09-2013 | minor revisions (reviews from Netherlands and Sweden) | Magdalena Homenko<br>Paweł Lańduch | GUS (Poland) |
| 0.4 | 30-10-2013 | minor revisions (Dutch review) | Magdalena Homenko<br>Paweł Lańduch | GUS (Poland) |
| 0.4.1 | 18-11-2013 | preliminary release | | |
| 0.4.2 | 11-03-2014 | minor revisions (EB review) | Paweł Lańduch | GUS (Poland) |
| 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 17:27 |