

QDDS - Documenting Survey Questionnaires throughout their Lifecycle

Oliver Hopt¹, Maximilian Stempfhuber¹

¹GESIS – Leibniz Institute for the Social Sciences

Email address of corresponding author: oliver.hopt@gesis.org

Abstract. Questionnaire development is a multi-step process. At every step, decisions are made. This article describes a novel software which logs these decisions rationale for subsequent questionnaire modifications. Therefore it provides a graphical user interface to create and manipulate questionnaires, and to track the changes of the individual steps of the questionnaire development process. The detailed change information adds to the comprehensiveness of change reasons and supports extended external documentation.

The documentation problem

Deciding to do a survey means committing oneself to work through a myriad of issues each of which is critical to the ultimate success of the survey. Designing and developing new instruments constitutes a complex and lengthy process. The development of a questionnaire involves dozens of revisions, in particular if the purpose is a methodologically demanding study (see for example Czaja & Blair 2005; Groves et al. 2004; Leeuw, Hox & Dillman, 2008; Presser & Blair 1994; Sirken, Herrmann & Schechter et al. 1999; Schnell, Hill & Esser 2008).

During the past 20 years, methods for testing and evaluating survey questionnaires have changed dramatically. In an effort to improve survey data quality, survey practitioners have significantly increased their use of an evolving set of questionnaire pre-testing methods (see for example Forsyth, Rothgeb & Willis, 2001; Presser, Cooper & Lessler et al. 2004; Tourangeau, R. 2003). The increasing use of cognitive psychology and methods derived thereof in the development of questionnaires for empirical social science research made single pre-tests obsolete. Due to the development of cognitive aspects of survey methodology, they have been superseded by a series of different pre-test methods.

There has also been a subtle shift in the goals of testing, from an exclusive focus on identifying and fixing overt problems experienced by interviewers and respondents during a pre-test to a much broader concern for improving overall data quality. Dozens of versions of the instruments are therefore generated and subsequently discarded during the development of a questionnaire, until a final version of phrasing of questions and layout of the questionnaire emerges.

Currently, there is a gap between the intellectual processes involved in developing a survey questionnaire and the tools used at different stages of the process, from the initial idea over pre-testing and to the final questionnaire which is then carried out. While every single version of a questionnaire inherently resembles the results and experiences from testing previous

versions, the reasons for these – sometimes subtle – changes are not reflected in the questionnaires themselves, but researchers have to use different tools and strategies for recording (motivations for) changes and to avoid problems due to the lack of thorough documentation:

1. Without detailed documentation, once the instrument is finalized, the rationale behind details in phrasing and format is no longer evident.
2. A manual change log, containing full documentation of individual changes and the underlying reasons results in an unwieldy, hardly manageable document which might be difficult to be used by other researchers beside the author.
3. The insight which prompted modification of a question, (category of) answer, format or sequence is almost inevitably lost because of the weak connection between the media used for a survey (e.g. Paper&Pencil, CATI) and the documentation of the development process (e.g. handwritten notes, word processing files).

There are many reasons to standardize change-tracking and documentation of the questionnaire development process. Uses for questionnaire documentation can be internal project documentation or external documentation to make research outcomes reusable by colleagues, and to enable them to evaluate or replicate one's own findings. Revisions of the questionnaire deserve being documented in detail for the sake of data analysis and archiving but more importantly with a view to the use of derivatives in future surveys, which requires professional knowledge management. Moreover, comprehensive documentation is essential for the teaching of empirical research methods.

To support these demands for permanent, detailed and comprehensive documentation of the questionnaire development process, the Questionnaire Development and Documentation System (QDDS) has been developed in a project funded by Deutsche Forschungsgemeinschaft (DFG). The goal of the project, based upon the ideas enunciated in Schnell, R., Kreuter, F. (2001), was to develop a software system which can be used to design different instruments typically used for questionnaires (e.g. Paper&Pencil, CATI etc.) and to model a wide range of question types and attributes (e.g. open and closed questions, life history calendar questions, answer defaults, integration of multimedia etc.). The complex internal representation used for modelling surveys is derived from the user's interaction with an easy to use graphical interface, which efficiently supports all actions involved in questionnaire design and manipulation. A graphical outline of the whole questionnaire gives an overview of its structure and selected aspects related to the individual questions. Questions can be arranged into a block which supports the management of related groups of questions but can also be used to present random questions to subjects: From within a block, a single question can be presented or all individual questions can be presented in random order.

From within QDDS, a number of different instruments can be created from a questionnaire definition, e.g. Paper&Pencil surveys and CATI programs. The number of different output formats is constantly growing. Therefore even a change of the used instrument during the design process of a survey will not mean to start with a blank sheet document but to use the existing questionnaire directly.

Concept of QDDS

After explaining that any loss of information is detrimental for survey research and survey creation processes in general, we have to resolve which information in detail is probably lost. In figure 1 this is illustrated on an abstract level. It shows that in usual practice only the final version of the questionnaire and the according codebook and data collection are kept in some kind of data archive. The loss of any of these documents means that the evolution of the questionnaire is no longer reproducible. If we keep all this data, a first step is done.

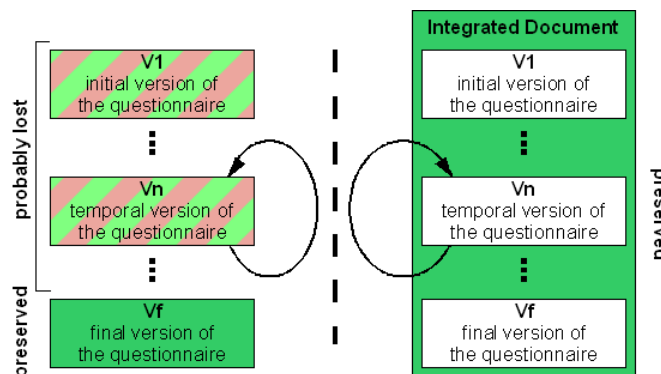


Figure 1: Theoretical workflow steps during questionnaire development

A closer look at the information available during the design process of a questionnaire shows that there are additional documents of interest, which will be lost mostly. These include, for example, all comments written down during pretest interviews or team meetings on how and why to change the questionnaire design. Moreover the data collected from pretest interviews will usually not be part of any archive. Figure 2 gives a snapshot on the collectable data and documents.

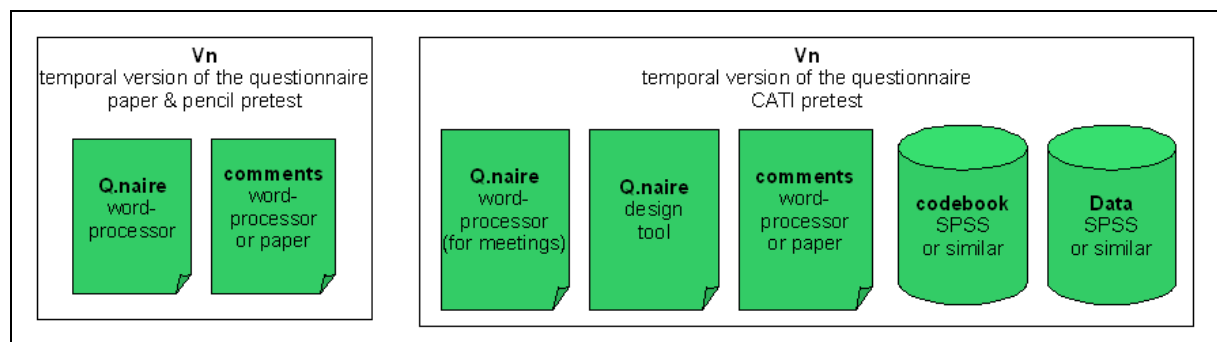


Figure 2: Overview on one iteration step during questionnaire development

An integrated documentation approach, as projected in QDDS II, will prevent most of these losses. Figure 3 shows a comparison of the complete amount of documents and information that are kept or probably lost within the entire process of a survey development in both the classic and the integrated approaches. It is obvious, that it is a problem to deal with all the heterogeneous documents from the actual development process. It implies that an overall documentation on every individual document and its content has to be implemented. To keep the resulting data from pretests is an individual task in any case. Besides it may be, that this

data is not of any interest for being not compatible with the data from the end version of the questionnaire.

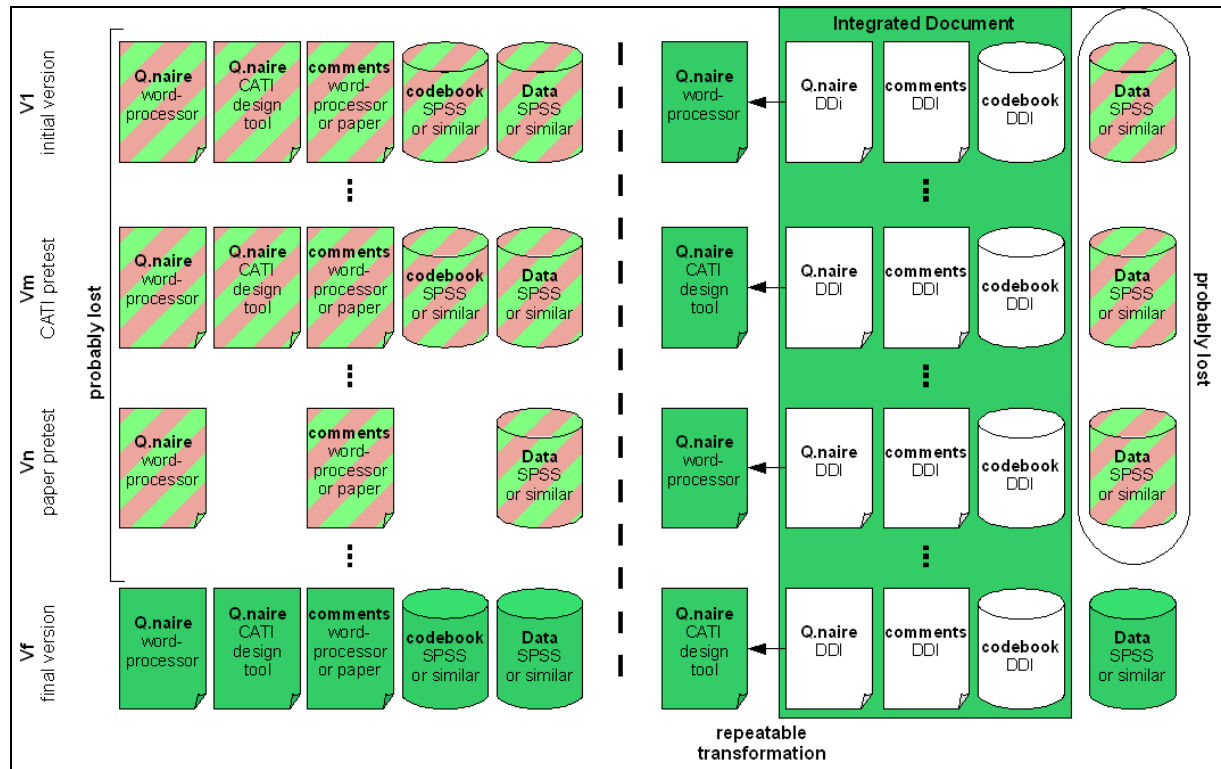


Figure 3: Complete overview on created documents during questionnaire development

Using QDDS

To develop a questionnaire with QDDS, the first step generally is to enter the general information on the new questionnaire like title, authors etc., the pre and post text for the communication with the interviewee and some project level metadata like funding agency, survey period or related literature.

In a second step, questions and blocks of questions will be added to build up the questionnaire core. Questions are meant to be items to be answered with an open statement/fact or with a selection from a set of answer defaults. For every question items, the user interface of QDDS offers entry forms containing the question type, a title, the question text, instructions, pre and post texts, answer defaults and filter definitions. Figure 4 shows the form to enter the main data for a question.

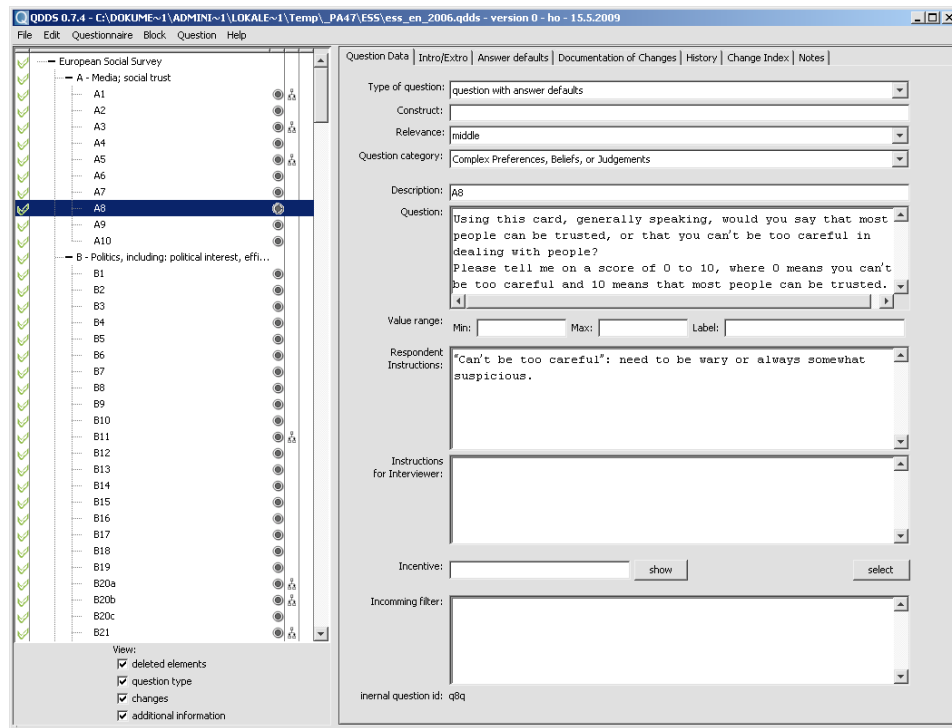


Figure 4. Screenshot of QDDS with form for question

The main purpose of blocks is to structure the questionnaire. Furthermore, they can contain information about repetitions and random control, and again pre and post texts. A special purpose of blocks is to define “block questions”, which in terms of QDDS are a number of questions with the same set of answer defaults displayed as a matrix (for example “the fear of being victim of several crimes in terms of much or less”).

Answer defaults can be set for questions of a closed type and for block questions. The form to enter the defaults will give one text line for each default. If there is a large number of defaults for one question, they may be structured by inserting gaps or headlines. Non-substantial answers are to be used by choosing them by category (answer denied, question not asked etc.) while the used text will be set generally on questionnaire level. If the question steers a filter, behind each default there will be a selection for the target question. Sets of answer defaults, once entered, can be exported for later re-use.

Another form of re-use within QDDS is to import questions and blocks from older questionnaires created with QDDS. These questionnaires can first be loaded in an overview and single elements can then be selected for importing in the active questionnaire. This will include all information except filter definitions. The import of complete questionnaires available in other formats depends on the specific format. Documents, which are “unstructured” in a formal meaning, like office documents, may not be imported for the reason of identifying single questionnaire elements. On the other hand, CATI programs which exist as ASCII files can be used as source documents. This already exists for Voxco and Nipo/Odin.

QDDS presents an overview of the questionnaire by displaying it as a tree, structured by blocks. Each block can be condensed to reduce the visible amount of questionnaire elements. Additionally, this tree is enriched by icons to display the state of questions: modification status, type of question, whether it is steering a filter, and existing external stimuli (like multimedia files).

Questionnaire can be exported to different file formats depending on the questionnaire mode (P&P, CATI etc.). The exported document will usually need some fine formatting work before use and exported CATI programs for dependent interviews will need programmers work to connect to the incoming data.

With the first export usually some kind of pre-test will be executed. To adopt the results of this pre-test, a new version of the questionnaire will be created in QDDS. For a new version, information about its reason and about the empirical base as part of the change documentation can be entered. General information about the major changes and their reasons can also be entered on the questionnaire level.

Each questionnaire element changed within a version can be annotated with comments on which parts have been changed (question text, answer defaults etc.), how they have changed (wording, filter etc.) and why the change was necessary. To overview the history of a questionnaire, there are two different approaches. First, the complete history of a single questionnaire element can be displayed containing all comments and the main textual components of the given element (e.g. title and question text). This view is presented in Figure 5. Second, an index of changes can be called to display all changes made to the entire questionnaire from a chosen version up to the actual version. This view will contain just short information on every changed element. Both views can be exported as HTML files to be re-used for external documentation or printing.

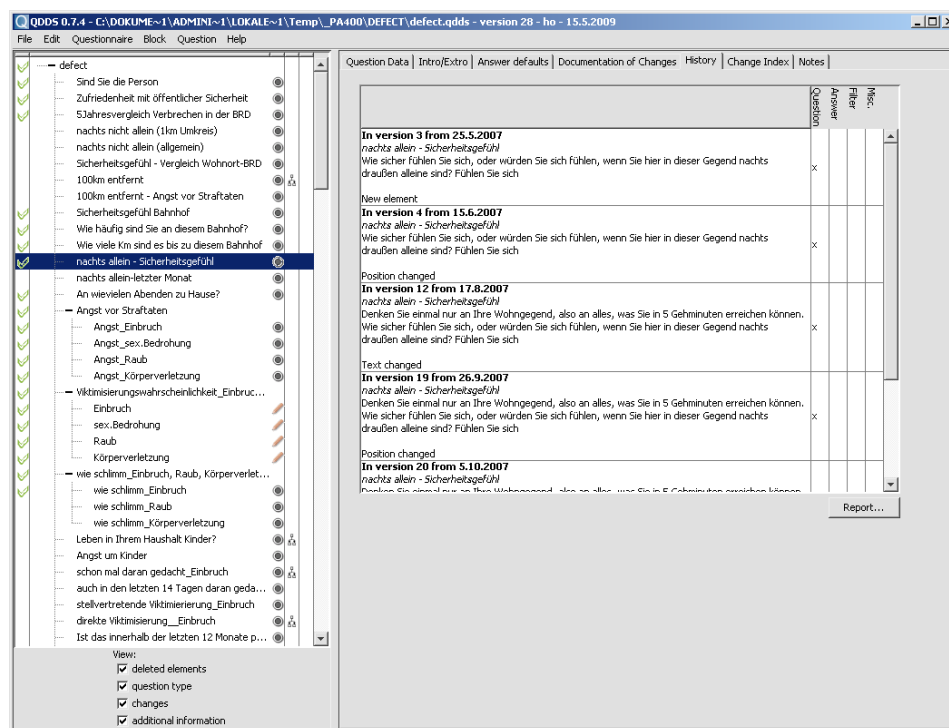


Figure 5. Screenshot of QDDS with the change history of a question

The QDDS software has already been used for documenting complex questionnaires, like ALLBUS 2004, SOEP 2006, ESS 2006 (Germany) and in the DEFECT research project, and experiences made in these contexts have been incorporated in the development process of QDDS.

Document format and handling

For many reasons the standard document format defined by the Data Documentation Initiative (DDI) is the best way to implement the approach of integrating questionnaires and change tracking. DDI was formed to create a standard format for study descriptions. Therefore a XML scheme was build up containing elements to describe the study in general and the raised variables. This format in its revision 2.1 is the basis for the documents created with QDDS II. The main modifications to this format are to insert some formatting information into the variable descriptions and to surround the structure with versioning information and general formatting properties. The main assumption for this was the equivalence of variable and question. This is in fact not like in real questionnaires, where often several variables are raised by one question item. But on the other hand, these question items - marked by just one question number or label - can be split into several questions raising just a single variable. Groups of variables are used for this special type called "block question" to document that they build up a combination of variables under one single question item. This requires, of course, some kind of rethinking for questionnaire builders.

Another challenge is to handle calculated variables. At the moment such variables are not part of QDDS II but are planned for future versions. They will be represented again by a special type of question and they will need a new entry field to document how they are calculated.

For using DDI, QDDS II produces files that can directly be used to be stored into data archives. On the other hand, XSL as a standard technology for transforming XML files can be used to create several output formats of questionnaires written in QDDS II. This can be different print formats (Open Document Format, LaTeX or RTF) or CAI systems. The output may differ depending on the questionnaire use (pretest, for interviewers, for interviewees etc.). Some of these are to be implemented during the current QDDS III project. To establish a new kind of export, just a new XSL stylesheet has to be committed to QDDS II.

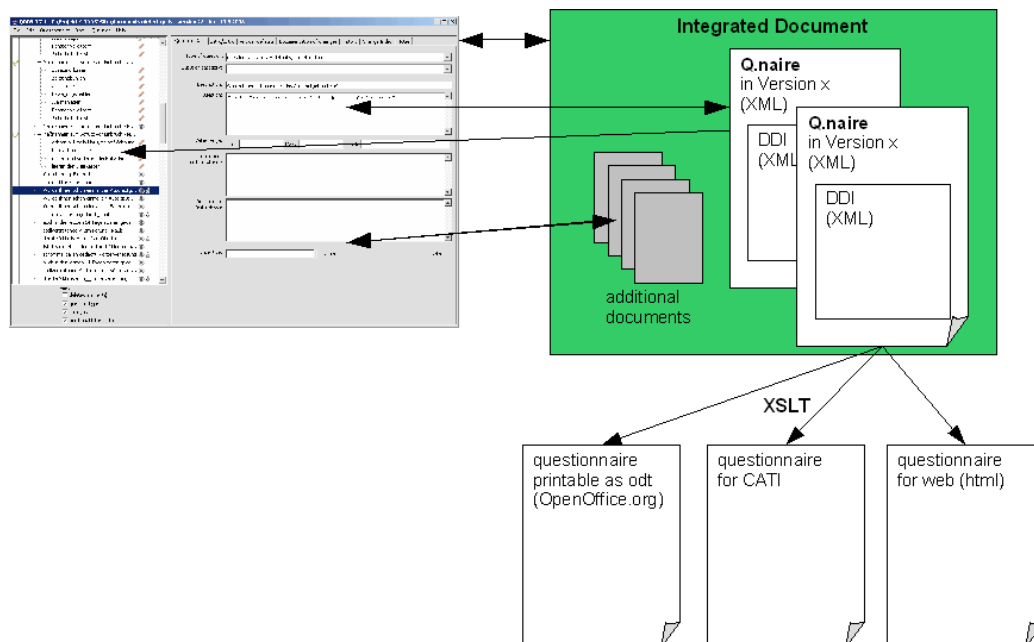


Figure 6. Information flow in and from documents

Currently a new revision of the DDI format (DDI 3.0) is available. QDDS II is prepared to handle the new format and can then support both DDI 2.1 and DDI 3.0 within the same application.

A QDDS II file is organized as a zip archive containing several files. There are XML files for each version of the questionnaire. Additionally there can be files for external excitation (like pictures, audio or video) and document templates for printing output (this approach is according to the Open Document Format). Based on the fact that all elements of a questionnaire including their specific properties are held in a structured document, these can be searched by various criteria like for example textual content or type of question.

Conclusion

The main new feature of the current version QDDS III, is the ability to document, store and retrieve questionnaires and survey meta-data for large survey archives. The program will e.g. allow complex searching, for surveys in which a battery of dichotomous questions is followed by an open question. Furthermore, it will show the variable name of the open question and the codes given to the answers. QDDS III will conform to the DDI-standard of data archives. Furthermore, every feature available in single purpose survey documentation systems like SOEPINFO will be implemented in QDDS III. Finally, about 50 important surveys will be documented within QDDS III to illustrate the use of the program. A first prototype can be seen on the project web site.

The software is available free of cost and will run on operating systems with a Java Runtime Environment (JRE) available. For Microsoft Windows, an installer package can be downloaded on the project web site www.qdds.org. To run the software on Linux or Unix platforms there is an archive containing all relevant parts. The user interface exists in English and German; additional translations are easy to implement.

References

- Czaja, R., Blair, J. (2005). 'Designing Surveys. A guide to decisions and procedures'. Thousand Oaks: Pine Forge Press.
- Forsyth, B., Rothgeb, J., Willis, G. (2001). 'Does question pretesting make a difference? An empirical test', in Presser, S., Couper, M.P., Lessler, J.T. et al. (eds.), *Methods for Testing and Evaluating Survey Questionnaires*, Hoboken, NJ: Wiley.
- Groves, R.M. et al. (2004). 'Survey Methodology', Hoboken, NJ: Wiley.
- Leeuw, E.D., Hox, J.J., Dillman, D.A. (2008). 'International Handbook of Survey Methodology'. New York: Psychology Press.
- Presser, S., Blair, J. (1994). 'Survey Pretesting: Do Different Methods Produce Different Results?' *Sociological Methodology* 24, 73-104.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Rothgeb, J. M., et al. (2004). 'Methods for Testing and Evaluating Survey Questions'. In S. Presser (Ed.), *Methods for testing and evaluating survey questionnaires: Wiley series in survey methodology*. Hoboken, NJ: Wiley-Interscience.

- Schnell, R., Kreuter, F. (2001). 'Neue Softwarewerkzeuge zur Dokumentation der Fragebogenentwicklung'. *ZA-Information*, 48, 56-70.
- Schnell, R., Hill, P., Esser, E. (2008). 'Methoden der empirischen Sozialforschung'. München: Oldenbourg. 8. Auflage
- Sirken, M.G., Herrmann, D.J., Schechter, S. et al. (eds.) (1999). 'Cognition and survey research'. NJ: Wiley.
- Tourangeau, R. (2003). 'Cognitive aspects of survey measurement and mismeasurement'. *International Journal of Public Opinion Research*, 15 (1), 3-7.