

The Missing Link: From Concepts to Questions in Economic Surveys¹

Diane K. Willimack
U.S. Census Bureau
Washington, D.C. 20233
diane.k.willimack@census.gov

Ger Snijkers
Statistics Netherlands
Heerlen, The Netherlands
g.snijkers@cbs.nl

1. Introduction

Data collected in business surveys typically rely on technical definitions and well-specified measurements. Moreover, the desired data are expected to be available in business records. However, underlying concepts may not be as clear-cut or as measurable as expected.

Questions in business surveys are intended to provide valid measurements of underlying economic concepts (i.e., construct validity) that often have many attributes. In survey practice, the variety of attributes may lead to mismatches with a respondent's interpretation or with available data, resulting in measurement error, as collected data fail to meet the intent of the survey question, the underlying concept, or the needs of data users. Survey designers are often unable to identify this ambiguity in the questions or concepts until cognitive pretesting or data collection are complete, demonstrating a missing link from concepts to questions and data that is often overlooked in business surveys.

The aim of this paper is to put a spotlight on this missing link, not only for survey questions but also for evaluating the validity of register data. In section 2, we will describe the steps in the process of constructing a questionnaire: conceptualization and operationalization in theory; followed by a discussion on practices in section 3. Section 4 discusses research methods that can be used to investigate concepts, ascertain attributes, identify measurements, and specify questions, to achieve construct validity at the design stage. Section 5 concludes this paper with a discussion on the use of these methods.

2. From theory to data: in theory

With a data collection process, data that are of interest to a researcher are collected. These data can be collected by conducting a survey or by using secondary sources, like registers. The data may be of interest for a number of reasons. The researcher is studying a theory, or wants to test some hypotheses. This approach is confirmatory research, which is often carried out at scientific organisations like universities. Or, the researcher may want to describe some phenomena. This approach is descriptive research, such as that carried out by national statistical bureaus. For both kinds of research the starting point is critical for getting the data the researcher wants. The starting point is to think carefully about the concepts that are to be studied (Willeboordse, 1998). The researcher first decides what concepts are of interest, and secondly, how to define those concepts.

Within a theory, the concepts of interest are described in a network of relationships among attributes. Attributes are the building blocks of concepts. An attribute is the smallest piece of measurable information that can be identified. A concept is defined by associating a concept to one or more attributes, and defining a specific algorithm with the attributes as input. A concept, C_1 , can be written as a function of attributes 1 to n (A_1, \dots, A_n), where the function definition denotes the algorithm:

$$C_1 = f(A_1, A_2, \dots, A_n)$$

¹This report is released to inform interested parties of research and to encourage discussion. Any views expressed on methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau or Statistics Netherlands. We would like to acknowledge Mattijn Morren of Statistics Netherlands, and Jennifer Whitaker, Eric Fink and Jennifer Beck of the Census Bureau for their helpful review comments.

Concepts can be simple, existing of only one attribute. In that case we have a one-dimensional concept, e.g. “gender.” In economic research, however, most concepts are multi-dimensional, indicating that a concept is composed out of more attributes.

Even a concept as seemingly straightforward as “employment” has different dimensions: Is the researcher interested in a straight “head count” or the number of full-time equivalents (FTEs)? What time frame or reference period should the respondent be considering – on a particular day, during a particular pay period, at the end of the month, or some sort of year-end average? Should part-time workers be included? What about temporary employees, leased employees, or contractors? Should employees on paid or unpaid leave be counted? And so on ... All these attributes of “employment” need to be identified. Taking these attributes together, and following a specific algorithm, defines the concept “employment”.

Similar reasoning can be applied to defining statistical units in economic studies. This concept also involves a number of attributes (see e.g., Ritzen, 2007; Willeboordse, 1998), like being active in the reference period, legal units, producing goods or services, having autonomy in making decisions, etc. The statistical unit may also depend on the topic of the survey (i.e., the concepts of interest). Again, the concept of “statistical unit” is defined by the identified attributes and follows a specific algorithm. In general, if attributes are ambiguously defined or missing, and if the relationship between attributes is ill-specified, a concept is ill-defined.

Attributes are the link to the data collection. In the case of a survey, questions in a questionnaire are specified to measure these attributes. Questions are the operationalization of these attributes; they are the operational definitions (Segers, 1977; De Groot, 1994) according to which the attributes are measured. This includes the wording of a question, answering options, and instructions. A question can be based on a single attribute, resulting in a one-dimensional question; it can also be based on a combination of attributes, resulting in a multi-dimensional question.

The next step is the design of the questionnaire, i.e. the construction of the measuring instrument. It involves: (re)writing the questions, (re)defining answering options, (re)writing instructions, composing blocks of questions that are related, putting the blocks and questions in order (e.g., according to theme), navigation, edit checks, as well as the visual design. Aspects of the design may vary depending on the data collection mode. Additionally, questionnaires may be tailored to subgroups.

Data are gathered by administering questionnaires with reporting units, resulting in the generation of a data file, in which each record represents a response (which may also be empty in case of unit non-response), and each column a variable. A variable can be related to one question; a variable can also be constructed out of two or more questions, resulting in a derived variable. The data file is used in the analysis to test the theory or to describe the phenomena we started with.

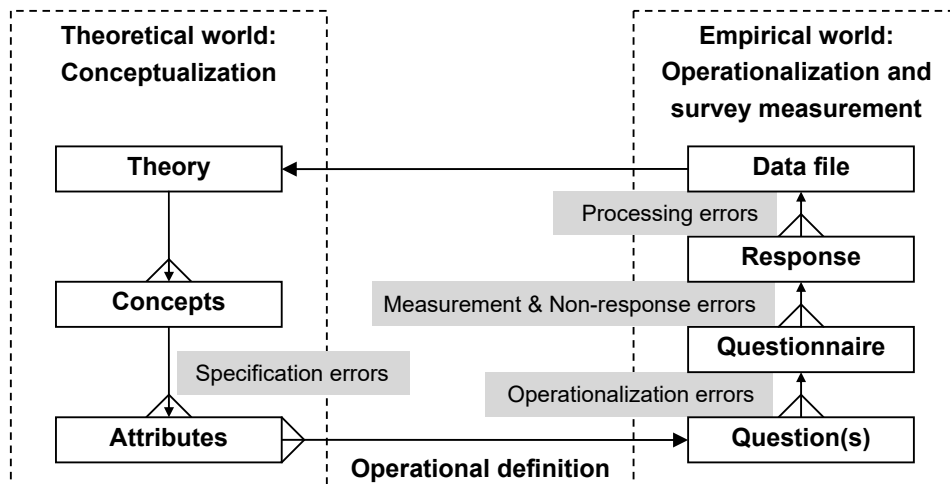
In the case where register data are used, no measuring instrument is needed. We already have a data file with variables. In that case we need to find out whether the variables in the data file are correlated with the attributes that have been identified, in order to construct the concepts of interest.

The steps in the process of going from theory to data are shown in Figure 1. Going from concepts to attributes is what we call conceptualization (or concept specification); going from attributes to questions and a questionnaire is called operationalization. If the conceptualization and operationalization are done properly, we get valid data with respect to construct validity (Groves et al., 2004) reflecting underlying economic concepts. Poor conceptualization leads to specification errors (Biemer & Cantor, 2007); likewise poor operationalization leads to operationalization or design errors (Snijkers, 2002).

To get valid data, construct validity is an important and necessary criterion. However, a questionnaire that is valid in the sense that it measures the underlying concepts (i.e., construct validity), may not be valid in the sense that the operationalization leads to other kinds of non-sampling errors (Snijkers, 2002; Lessler & Kalsbeek, 1992; Groves, 1989), such as item non-response or measurement errors, also shown in Figure 1. Rounding out the error profile are processing errors that occur as responses are coded and captured to form a data file (Groves et al., 2004).

In this paper we will focus on the construct validity of data collected using survey questionnaires or obtained from registers.

Figure 1. From theory to data



3. From theory to data: in practice

The process of conceptualization and operationalization is not an easy one. It is about bridging the gap between theory and measurement. According to Schwarz (1997), the processes of conceptualization and operationalization are only partly covered in the literature on survey methodology. A lot of attention is given to the operationalization of questionnaires, but the step that precedes it, concept specification, is not well discussed. We feel this is still the case today, and holds true even more for business surveys. In the practice of business surveys we come across many questionnaires in which the gap between theory and measurement is not well bridged. To quote Schwarz (ibid., p. 30): "... the research objectives of many studies are surprisingly ill-defined. Asking a researcher what exactly should be measured by a question for which purpose frequently elicits vague answers – if not different answers from different researchers involved in the same project."

Even after 15 years this has not changed much. Questionnaire developers and statisticians are often told by researchers:

- "These questions have been done in this way for years and we have always gotten data, so evidently respondents understand the questions and are able to report the data."
- "Changing the questionnaire would cause disturbances in the time series."
- "The questions have to be asked in a particular way because of Eurostat or government regulations, or because they are the outcome of negotiations with the stakeholders and data users."
- "Only those respondents who have or do the item of interest will understand the technical language and answer the associated question, while all others will correctly skip over the question."

3.1. Consequences of inattention to concept specification

The lack of attention to clear specification of the underlying concept results in questionnaires that are hard for businesses to interpret and complete, even if the questionnaires by themselves are well-designed. One could say that in these cases the concept specification is carried out by the respondents, with variation across respondents as different respondents define the concept differently. This obviously has a negative effect on the data quality.

It is our experience that an explicit step of identifying attributes (as is illustrated in Figure 1) is often missing in economic research (conducted by national statistical institutes). One goes directly from concepts to questions, resulting in specification errors. This has a number of consequences with regard to data quality:

- Missing variables: Researchers may find that variables are missing when they start analysing the data. Since the concepts are not defined properly by identifying the related attributes, some attributes may not have been specified. For these attributes, no questions have been defined. As a consequence, parts of the concepts of interest have not been measured. Register data used by researchers may also suffer from unmeasured attributes.

- Effects on variability or bias: If the concepts are vague and ambiguous, the questions in a questionnaire will also be vague and ambiguous. The questions are specified without knowing what they actually measure, resulting in mismatches with a respondent's interpretation or with available data, i.e. comprehension and retrieval problems per the cognitive response model (Tourangueau, Rips & Rasinski, 2000). For these questions, respondents are forced to determine what the question is asking for on their own.

Cognitive research has shown that respondents tend to interpret or re-interpret ill-specified questions relative to their own contexts (Clark & Schober, 1992), and provide answers based on data that are readily available. This process also occurs among business respondents as well (Geisen, 2007; Willimack, 2008). As different respondents answer the same question using differing contexts, the result is unforeseen, unknown and unmeasured measurement errors. Undetermined effects on the observed sample variance of survey estimates result from the increased variability in response strategies, or bias may result from systematic response error.

In case of the use of registers, ill-specified concepts may result in the use of register variables that are not well correlated with the concepts of interest. Register data may also suffer from specification errors and measurement variation.

- Non-response: If the effort to complete a questionnaire becomes too large because of comprehension and retrieval problems due to poor operationalization of the questionnaire, respondents may decide to leave questions unanswered (item non-response), or abandon the questionnaire altogether (unit non-response).
- Response burden and trust: Comprehension and retrieval problems in the field also increase the response burden, as diligent respondents expend extra effort trying to figure out what data items a question is asking for relative to data available in their business records. There may also be a negative effect on trust in the survey organisation and the survey data, as business respondents doubt the knowledge and expertise of those that collect the data and publish the results.

3.2. The misunderstood role of cognitive testing

With the advent and acceptance of cognitive research methods, many researchers, stakeholders and survey managers have come to consider cognitive testing to be the “magic bullet” for resolving *any* issues with specifying survey questions. Results are expected to provide considerable aid in identifying and clarifying underlying conceptual attributes and their measurements. However, cognitive testing begins with survey questions that have already been drafted, presumably to measure well-specified attributes. Its purpose, then, is to evaluate the effectiveness of survey questions in eliciting answers from respondents that meet the question's intent as operationalized (see Snijkers, 2002, and Willis, 2005, for more details).

Survey methodologists often start out with a set of draft questions formulated around “observable variables” believed by researchers and stakeholders to be associated with the underlying concepts, only to find during testing with respondents that these supposedly observable measurements are inadequate, inappropriate, unfamiliar, or even incorrect, from respondents' perspectives. In economic surveys, underlying concepts based on economic theory may bear little resemblance to practical definitions and measurements, and the associated data, that businesses track for management and regulatory purposes (Willimack & Nichols, 2010).

It is not particularly unusual for cognitive interviews to end up exploring the conceptual underpinnings of survey questions when it becomes apparent that they fail to meet their intended purpose. Thus, when activities for conceptualization and operationalization have not been systematically undertaken prior to the questionnaire development stage, cognitive interviews often devolve into a hybrid of cognitive probes and exploratory questions (Willimack, 2008).

We find this backtracking to be all too frequent in the development and pretesting of establishment survey questionnaires. As a result, respondents are frustrated by the unclear, seemingly odd, questions being pretested. Researchers and stakeholders are annoyed that the cognitive interviews failed to provide potential solutions to the problems identified during the testing. Then time is wasted as this backtracking requires re-formulating questions that may still prove to be inadequate in subsequent testing if attributes remain poorly specified and measurements ill-defined.

We contend that inefficiencies experienced during questionnaire development due to backtracking may be avoided through attention to conceptualization and operationalization. Some methods will be discussed in the next section.

4. Methods to be used in the conceptualization and operationalization

A number of methods can be used to investigate concepts, ascertain attributes, identify measurements, and specify questions to achieve construct validity at the design stage. These methods are: Dimension/Attribute Analysis, Content Development Techniques, Questionnaire Mining, Early-stage Scoping Interviews / Exploratory Focus Groups, Concept Mapping, Feasibility Studies, Accounting Expert Reviews, and Record-keeping Studies.

Hox (1997) distinguishes between top-down, theory-driven and bottom-up, data-driven approaches. The theory-driven approach starts at the left side of Figure 1, with constructs and works towards observable variables; the data-driven approach encompasses the right side, with operationalizations and observations, and works towards theoretical concepts. We will follow this dichotomy of methods.

4.1. Theory-driven approaches

- **Dimension/Attribute Analysis (Hox, 1997):**
Hox calls this method Dimension/Indicator Analysis, where an indicator is similar to what we've called an attribute. According to Hox, many researchers probably view dimension/indicator analysis as *the* approach to bridge the gap between theory and measurement. This analysis basically follows Figure 1: Empirical attributes (or indicators, according to Hox) are specified for the concepts in a theory. The process of concept specification is driven by an existing theory, logical reasoning, or may also be based on results from previous research. The result is a network of concepts that are logically tied together, and taken all together make up the theory. Next, appropriate empirical attributes are defined for the concepts. The process of concept specification ends when one or more indicators can be identified for all concepts. The attributes are the basis for the questions.
- **Content Development Techniques (Willimack, et al., 2004; Mulrow et al., 2007a):**
A number of techniques involve data users, researchers and stakeholders, in concert with survey methodologists, in developing survey content by specifying and clarifying data needs. Workshops may be conducted with data users to 1) identify how and why the data of interest are to be used; 2) identify gaps in existing data; 3) gain insight into data needs for specific issues; and 4) create a preliminary set of data priorities. Draft lists of requested data items may be iteratively ranked by different groups of stakeholders, experts, and trade representatives, while survey personnel indicate items previously collected with high quality. Panels of industry experts may be convened to aid researchers in drilling down from concepts to attributes to common definitions and metrics. These and other similar methods aid development of early drafts of questions and questionnaires for additional examination using data-driven approaches involving respondents.

4.2. Data-driven approaches

The following methods involve data provided by respondents, or information gleaned from their perspectives, which are typically qualitative in nature. These methods reflect varying degrees of exploration as concepts and their attributes gain specification.

- **Questionnaire Mining:**
In Questionnaire Mining, the researcher starts with existing questionnaires, from which questions are selected that seem to be relevant in the context of the study. The questions are reviewed as to how they are related to the central concept(s), i.e. a meta-data analysis. The reviewing is done in close collaboration with content matter experts, since this requires subject matter knowledge, resulting in an expert appraisal of the sampled questions. Reports of pre-test studies for these questions, analysis of paradata, and item non-response, examination of the types and frequency of edit failures associated with these questions, or the data analysis itself, can be used to study the validity and measurement problems of these questions. The result is a selection of questions that may be used in subsequent measuring instruments, in the same or modified operationalizations. Since this approach starts with existing operationalizations, it is essential not to miss any important aspect of the research domain, such as attributes that are not covered in the existing questionnaires.

- **Early-stage Scoping Interviews / Exploratory Focus Groups:**
If it is at all uncertain how concepts can be measured, in-depth exploratory interviews can be conducted with a small number of representatives of the target population. This may be accomplished through one-on-one, early-stage scoping interviews (Stettler & Featherston, 2010) or by using focus groups (Snijkers, 2002). Participants are asked how they interpret the concepts, how they define and structure attributes of the concept relative to their business activity, and what is or is not included. Also we learn about terms respondents use in relationship to the concept, so that these terms can be used in the questionnaire to ensure they are interpreted as intended. These interviews also give an idea about the availability of the data: are the requested data easy to collect or is it a burdensome process? In addition, we learn the identity or positions of these employees within businesses who have to be contacted to get survey participation and who have access to the desired information (the informant). Interview or focus group participants are selected in such a way that a wide variety of views is collected. The discussions are recorded, or notes are taken, and findings are reported to project sponsors and stakeholders.
- **Concept Mapping:**
A more structured way of studying the interpretation of concepts is by applying concept mapping in focus groups (Hox, 1997). This process involves six steps. In the first step, the subject of the focus group is specified, and participants are selected from the target population. The second step, “statement generation”, is a brainstorm session with the participants, to generate statements that describe many relevant aspects (i.e., attributes) of the concepts under study. These statements are printed; and each participant gets a pile with statements. In step three, the “structuring” step, the participants individually sort the statements into different piles according to their own views. Each pile contains related statements. The piles are combined in a group similarity matrix. This matrix is analysed in step four “statement representation”, by a multi-dimensional scaling technique. The result is a concept map, showing clusters of the statements. These clusters may be interpreted as the attributes that are related to the studied concepts. In the next step, “concept map interpretation”, this map is discussed in the focus group. The participants discuss possible meanings and acceptable labels for each statement cluster. The final step, “utilization”, concerns the translation of the statements into survey questions. The concept map gives guidance to structuring the questionnaire into blocks of related questions. Haraldsen (2003) applied this method to operationalize the concept “perceived response burden,” aimed at developing a questionnaire to measure this concept.
- **Feasibility Studies (Willimack et al., 2004):**
In a feasibility study, a small number of respondents visited on site are asked about the information that will be requested in a survey collection: is the data available, and easy to retrieve? Survey personnel conduct meetings with business representatives who are involved in the data collection process. A pre-specified topic list or agenda is followed to discuss the concepts of interest, the definitions, and the availability of the required data. Information collected during these visits helps determine whether concepts are measurable, what the questions should be, how to structure the questions and questionnaire, and to whom the questionnaire should be sent. This method does not yet require a fully developed questionnaire, but in contrast to the methods above, it does require specification of the concepts of interest prior to applying it.
- **Accounting Expert Reviews (Willimack & Gibson, 2010):**
Professional accountants aid concept clarification and definition as “subject matter experts” in business record-keeping and accounting practices. Informed by professional standards, regulatory bodies and law, professional accountants help to associate likely actual measurements to concepts, through their knowledge of accounting databases and interfaces (e.g., human resources, production management) where data elements may be an input to or product of accounting processes. During questionnaire development, accounting experts may work with survey sponsors, stakeholders, and content specialists, helping to evaluate correspondence between underlying concepts and business terms and constructs, along with the reportability and quality of the desired data. Accounting experts may also assist with defining concepts through comparisons with public information, performing analytical reviews of data elements (comparative ratios, relationship between data items, industry practices), and evaluating expected data sources and compilation methods to determine whether they are likely to produce the desired data. This method results in suggestions for question wordings and specifications, which may in turn be investigated via record-keeping studies.
- **Record-Keeping Studies (Willimack et al., 2004; Mulrow, et al., 2007b):**
In a record keeping study, a small number of cases are selected from the target population to be consulted about the data they have available in their business records. This method studies the availability of the

information, how it is structured within the business' records, and what the underlying definitions and business purposes are for the data. Thus, we find out what data may be collectable and what they consist of, in order to gauge the degree to which recorded and available data align or fail to align with attributes of the central concept(s). Results are used to assess the collectability of data adequate to measure the desired concepts, and to aid in writing effective survey questions for obtaining these data.

5. Discussion: an iterative approach

In practice we advocate not to use only one approach, but to combine methods in an iterative manner. Following Dimension/Attribute Analysis and Content Development Techniques, preliminary drafts of the questions can be developed. However, before starting the operationalization, data-driven approaches, like Questionnaire Mining, Early-stage Scoping, Exploratory Focus Groups and Concept Mapping can be used. Once concepts and attributes have been studied using these methods, next a Feasibility Study can be carried out. Finally, during the design of the questions, Accounting Expert Reviews and Record-keeping Studies can be applied. After a proper conceptualization and operationalization, a questionnaire can be fully drafted, and then pre-tested using techniques such as cognitive interviewing (see Snijkers, 2002; Willimack et al., 2004; Willis, 2005).

The following example may illustrate what can happen when a questionnaire is pre-tested without clear concept specification.

Recently, Statistics Netherlands (SN) participated in the pre-test of a newly designed questionnaire that was aimed at measuring the Global Value Chains (GVC) in businesses (Morren & Snijkers, 2011a, 2011b). GVC is a European project which attempts to develop methodologies and measurement methods to analyse globalisation practices of businesses and their impact on European economy and employment in the period from 2007 through 2011. The GVC questionnaire includes questions about the company in general, like the number of personnel involved, and the global value chain and sourcing activities. e.g., the countries involved, motivations for international and back sourcing, and future sourcing plans and barriers expected herein. The questionnaire will be sent to large companies, with 100 or more employees, from various sectors of industry and services. It is planned to be implemented in 2012.

Prior to the development of the questionnaire, in the conceptualization stage, a lot of discussions took place among subject matter experts with regard to the focus of the study, including the concepts of interest and what should be included in the questionnaire. Because of time pressure, a questionnaire was drafted prior to completion of this stage. In the pre-test, SN visited ten businesses to study the comprehension of this questionnaire and the availability of the data. Even though the results of the pre-test indicated that the questionnaire could be answered and the data were available in the business records, discussions with respondents ended with issues like: "What is this question supposed to measure?" It was hard to come up with recommendations to improve the questionnaire, since underlying concepts and attributes were not well-specified, and consequently an anchor or clear reference for the questions was missing. This lack of reference was reported back to researchers in charge. The next step in the design of this survey is a more detailed discussion between the researchers and the survey methodologists on defining the concepts and associated attributes that are to be measured.

From this example we can see an iterative process emerging. From the start careful attention was given to the conceptualization and operationalization: the project started with the specification of concepts using Content Development Techniques. Also careful attention was given to pre-testing the questionnaire. However, the questionnaire was drafted and pre-tested too soon. This in turn had to be compensated for by reviewing the questionnaire in a second stage. As a consequence, a new questionnaire needs to be pre-tested again. The starting point was a theory-driven approach, then one moved to the empirical world, and back again to theory.

We propose that in general the process of conceptualization and operationalization be iterative. This process relies on and benefits from using a combination of top-down and bottom-up approaches as discussed in this paper, involving theoretical researchers, stakeholders, questionnaire developers, survey designers, and respondents at an early stage in the process. If this is done properly, at the end of the process, both concepts and questionnaires are well defined by associated attributes and their operational definitions (i.e. questions).

To achieve this, we recommend these activities be explicitly included in the survey plan, to ensure adequate time to draft, revise, and pretest the questionnaire before using it in the field. If this is not well planned, and a questionnaire is drafted too soon, time may be too short to develop a valid questionnaire, resulting in undesirable consequences for data quality and respondent burden.

6. References

- Biemer, P., and D. Cantor, 2007, Introduction to Survey Methods for Business and Organizations. Short course, 18 June 2007. Presented the 3rd International Conference on Establishment Surveys, 18-21 June 2007, Montreal, Canada
- Clark, H., and M. Schober, 1992, Asking Questions and Influencing Answers. In: Tanur, J.M. (ed.), Questions about Questions: Inquiries into the Cognitive Bases of Surveys (pp. 15-48). Russel Sage Foundation, New York.
- De Groot, A.D., 1994, Methodology (In Dutch: Methodologie: Grondslagen van onderzoek en denken in de gedragswetenschappen). Van Gorcum, Assen (http://www.dbnl.org/tekst/groo004meth01_01/).
- Giesen, D., 2007, The Response Process model as a Tool for Evaluating Business Surveys. Proceedings of the 3rd International Conference on Establishment Surveys (pp. 871-880), 18-21 June 2007, Montreal, Canada. American Statistical Association, Alexandria.
- Groves, R.M. 1989, Survey Errors and Survey Costs. Wiley, New York.
- Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau, 2004, Survey Methodology. Wiley, Hoboken.
- Haraldsen, G., 2003, Searching for Response Burden in Focus Groups with Business Respondents. In: Prüfer, P., M. Rexroth, and F. Fowler (eds), Proceedings of the 4th Workshop on Questionnaire Evaluation Standards (pp. 113-123), QUEST 2003, 21-23 October 2003, Mannheim, Germany. ZUMA Nachrichten, Spezial Band 9, Mannheim, Germany.
- Hox, J., 1997, From Theoretical Concepts to Survey Questions. In: Lyberg L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), Survey Measurement and Process Quality (pp. 47-69). Wiley, New York.
- Lessler, J.T., and W.D. Kalsbeek, 1992, Nonsampling error in surveys. Wiley, New York.
- Morren, M., and G. Snijkers, 2011a, Test Plan of the Global Value Chains (GVC) Questionnaire. Report Statistics Netherlands, Heerlen, The Netherlands.
- Morren, M., and G. Snijkers, 2011b, The Global Value Chains Questionnaire: Results of the Dutch field test. Report Statistics Netherlands, Heerlen, The Netherlands.
- Mulrow, J.M., L. Carlson, J. Jankowski, B. Shackelford, and R. Wolfe, 2007a, A Face-Lift or Reconstructive Surgery? What Does It Take to Renew a 53-Year Old Survey? Proceedings of the 3rd International Conference on Establishment Surveys (pp. 208-213), 18-21 June 2007, Montreal, Canada. American Statistical Association, Alexandria.
- Mulrow, J.M., S. Freedman, and R. Rutchik, 2007b, Record Keeping Studies – Love ‘Em or Leave ‘Em. Proceedings of the 3rd International Conference on Establishment Surveys (pp. 56-62), 18-21 June 2007, Montreal, Canada. American Statistical Association, Alexandria.
- Ritzen, J., 2007, Statistical Business Register; Content, Place, and Role Economic Statistics. Proceedings of the 3rd International Conference on Establishment Surveys (pp. 179-191), 18-21 June 2007, Montreal, Canada. American Statistical Association, Alexandria.
- Schwarz, N., 1997, Questionnaire Design: The Rocky Road from Concepts to Answers. In: Lyberg L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), Survey Measurement and Process Quality (pp. 29-45). Wiley, New York.
- Segers, J.H.G., 1977, Sociological research methods (in Dutch: Sociologische onderzoeksmethoden). Van Gorcum, Assen/Amsterdam.

- Snijkers, G., 2002, Cognitive Laboratory Experiences. On Pre-testing Computerised Questionnaires and Data Quality. Ph.D. Thesis, Utrecht University and Statistics Netherlands, Heerlen.
- Stettler, K. and F. Featherston, 2010, Early Stage Scoping: Building High Quality Survey Instruments Without High Costs. Presented at the 2010 Annual Conference of the American Association for Public Opinion Research, Chicago.
- Tourangeau, R., L.J. Rips, and K. Rasinski, 2000, The Psychology of Survey Response. Cambridge University Press, Cambridge.
- Willeboordse, A. (ed.), 1998, Handbook on the Design and Implementation of Business Surveys. Eurostat, Luxembourg.
- 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. & P. Gibson, 2010, Collaborating for Quality: A Cross-Discipline Approach to Questionnaire Content Evaluation in Business Surveys. Proceedings of the 2010 European Conference on Quality in Official Statistics, Helsinki, Finland.
- Willimack, D.K. & E.M. Nichols, 2010, A Hybrid Response Process Model for Business Surveys. Journal of Official Statistics.
- Willimack, D.K., L. Lyberg, J. Martin, L. Japac, and P. Whitridge, 2004, Evolution and Adaption of Questionnaire Development, Evaluation, and Testing Methods for Establishment Surveys. In: Presser, S., et al. (eds), Methods for Testing and Evaluating Survey Questionnaires (pp. 385-407). Wiley, Hoboken.
- Willis, G.B., 2005, Cognitive Interviewing. A Tool for Improving Questionnaire Design. Sage, Thousand Oaks.