VERSITA

Journal of Official Statistics, Vol. 30, No. 1, 2014, pp. 163–166, http://dx.doi.org/10.2478/jos-2014-0009

Book Review

Books for review are to be sent to the Book Review Editor Jaki S. McCarthy, USDA/NASS, Research and Development Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.
Email: jaki_mccarthy@nass.usda.gov

Jennifer Madans, Kirsten Miller, Aaron Maitland, and Gordon Willis (Eds). *Question Evaluation Methods: Contributing to the Science of Data Quality*. Hoboken, NJ: John Wiley & Sons, Inc. 378 pp. 2011. Paperback: ISBN 9781118037003, price USD 64.20. E-pub: ISBN 9781118036983, price USD 52.99.

Website Q-Bank: http://wwwn.cdc.gov/qbank/

This book grew out of an interdisciplinary workshop on question evaluation methods and has as its goal to bring together knowledge from leading experts across different methods. The book consists of seven sections, or rather seven extended chapters, as each section contains a primary chapter describing a specific method and one or two shorter discussion chapters.

The first section opens with an excellent overview by Jack Fowler on behavior coding as a tool for evaluating survey questions. This chapter describes how behavior coding of interviewers and respondents is done and presents empirical evidence of its significance. It concludes with an outline of how this method should be fitted into question evaluation protocols and presents a well chosen reference list including key references in this field. The Fowler chapter provides an introduction to the novice in behavior coding and a good summary for those who have some experience with question evaluation. The two response chapters are aimed at the more advanced researcher. Nora Cate Schaeffer and Jennifer Dykema present a conceptual framework on how the interaction between respondent and interviewer affects data quality. They also present two very interesting summary tables (Tables 3.1 and 3.2) on the empirical associations between interviewers' and respondents' behavior and measurement quality. They then introduce the reader to conversation analysis as a tool and illustrate this with excerpts from the Wisconsin longitudinal study. Alisu Schoua-Glusberg broadens the discussion and focuses on the sociocultural context in which the survey interview takes place. Her remarks concern behavior coding as an evaluation tool, but are equally worthwhile for other question evaluation methods, such as cognitive interviews. As international, cross-cultural, and multilingual studies take on a greater importance in the modern world (cf. Harkness et al. 2010), researchers should realize that respondents will have different degrees of familiarity with the survey process;

in developing and pretesting survey questions, differences in communication styles in different cultural groups should be taken into account.

The second section on cognitive interviewing opens with a chapter by Kristen Miller, who gives a theoretical review of development of cognitive interviewing and describes a new integrative paradigm for question development and testing. This chapter is clearly intended for the knowledgeable and methodologically interested reader and is not meant as an introduction to cognitive interviewing. Those seeking such an introduction should read Willis's (2005) book first. The following two chapters by Gordon Willis and by Fred Conrad are critical rejoinders, and like the Miller chapter their discussion is aimed at cognitive survey methodologists.

Question evaluation and cognitive survey methodology are often seen as more qualitative approaches, but the next sections prove that this view is incorrect. Statistical modeling provides us with powerful tools for investigating measurement error and evaluating questionnaires. These quantitative methods are used in phase two of questionnaire evaluation. In phase one, the questionnaire is developed and pretested using more qualitative approaches, for example, expert evaluation and cognitive interviewing. The improved questionnaire is then implemented in an actual survey, ideally a field test.

In their chapter (Section 7) Brian Harris-Kojetin and James Dahlhamer describe what field tests are and the importance of collecting additional data, such as interviewer feedback. They illustrate this with examples from US federal statistical surveys. Field tests are fairly common in daily survey practice. Less common are the use of Multi-Trait Multi-Method (MTMM) matrices and specific experiments to collect data for quantitative questionnaire evaluation. Section five is devoted to split-sample experiments as a tool for collecting data for question evaluation. Jon Krosnick starts with a brief review of the experimental method and provides several insightful examples of methodological experiments on question wording, formats, and context. These are relatively large-scale field experiments aimed at quantitative analysis. Johnny Blair adds to this an outline for a more qualitative cognitive interview experiment. Theresa de Maio and Stephanie Wilson expand on this by emphasizing the importance of integrating a qualitative and quantitative approach. To quote: "this mixed-method approach allows us to understand what survey questions are actually measuring, and make better decisions about which questions to field".

Section six on the multitrait-multimethod approach deals with a special kind of experimental setup and its analysis. In his introductory chapter, Duane Alwin first describes the MTMM design as an approach to systematically collect data and gives a historical overview starting with the work of Campbell and Fiske in psychology and of Andrews in sociology and survey research. The concepts of reliability and validity in MTMM and in classical test theory are clearly explained and trait validity versus construct validity is discussed. Special attention is given to the role of memory in MTMM designs and recent applications of the MTMM approach. This chapter is a mixture between data collection and data analysis; data collected according to an MTMM design are by default analyzed using a Structural Equation Modeling (SEM) approach. In his response to Alwin, Peter Mohler gives an extended example of an MTMM study from the European Social Survey.

After the collection of large-scale quantitative data, be it through a regular survey, a specially designed field test, or a specific (experimental) design, there are various

statistical methods that provide powerful tools for quantitative questionnaire evaluation. For instance, the previously mentioned SEM approach can also be used to carry out multigroup comparisons and investigate measurement equivalence across different cultural or national groups (Vandenburg and Lance 2000; Hox et al. 2010). In cases where multiple items are used to measure one well-defined construct, Item Response Theory (IRT) is a promising analysis tool. Section three opens with a brief but informative chapter on Item Response Theory (IRT) and how it can be applied to questionnaire evaluation. IRT modeling focuses on scales that measure an underlying construct, using multiple items and a strict psychometric model. Bryce B. Reeve introduces the principle of IRT and illustrates how it can be applied in evaluating and refining questionnaires. He then introduces Computer Adaptive Testing (CAT) where a combination of qualitative pretest methods, such as expert evaluation and cognitive testing, and quantitative data analysis is used to produce an item bank with IRT-calibrated items. In CAT a respondent is then presented with an item in the middle range and an estimate is made of the person's scale score based on the response; then another item based on this estimate is selected from the item bank, and the process is repeated until the desired precision is reached. CAT allows for short questionnaires, adapted to the person's ability, with the desired precision. This is illustrated with PROMISS (Patient-Reported Outcomes Measurement Information System). In his rejoinder, Ronald Hays provides the reader with additional examples of the use of IRT in question evaluation. He also emphasizes that IRT analysis is extremely useful for detecting problematic items and building libraries of well-performing items, but that qualitative methods are needed to understand why an item performs badly. The next rejoinder by Clyde Tucker et al. is less a discussion of IRT and more an introduction to Latent Class Analysis (LCA) as a tool for questionnaire evaluation. In LCA an attempt is made to find an underlying latent categorical nominal or ordinal variable (latent classes) that explains the relationship between a number of observed variables. This is well illustrated with an example where LCA is used to classify respondents into good, fair, and poor reporters of expenditures.

Latent Class analysis (LCA) is then further introduced in section four by Paul Biemer and Berzofsky. In their conclusion they state that LCA is challenging for a novice. Their chapter proves them right; it requires more advanced statistical knowledge than the other chapters. Biemer and Berzofsky present the reader with a statistical introduction to LCA, its assumptions, and how to handle some common statistical problems. Together with the examples in the previous chapter it gives a good impression of how LCA can be used in questionnaire evaluation and in discovering response tendencies. In her rejoinder, Frauke Kreuter summarizes a comparison of different traditional questionnaire testing techniques (e.g., expert evaluation) and LCA; she also offers good guidelines on how test material and analysis results should be incorporated into question banks. Finally, Janet Harkness and Timothy Johnson go beyond LCA analysis as such, addressing issues in question design and pretesting that are somewhat neglected in general discussion, such as context effects.

Reasons to buy this book: Renowned experts from different disciplines introduce and discuss qualitative and quantitative methods of questionnaire evaluation. The methods introduced go beyond standard question evaluation methods such as expert evaluation and cognitive interviewing and focus on the collection and analysis of quantitative data for questionnaire evaluation. The quality of the contributions is high. The book is

accompanied by the very worthwhile Q-bank website (http://wwwn.cdc.gov/qbank/). Q-bank goes beyond more traditional question banks, providing the reader with an online database of questions that have been evaluated as well as their accompanying question evaluation reports.

Reasons not to buy this book: Although the book aims at a wide audience, not all chapters are easily accessible. Due to the format, a large introductory chapter followed by shorter rejoinders, the discussion aims at experts in the field. It is the well-edited proceedings of a multidisciplinary workshop and still reads as such.

In sum: I am glad I have read the book. I will certainly use (parts of) it in teaching advanced courses in survey methodology and it is a good accompaniment to the well-known earlier book by Presser et al. (2004). Both books should be in the library of survey researchers and statisticians in the private sector, government and academia, and the library of my institute now has both. However, if you have to advise a master or graduate student with limited monetary resources and have to choose one, I would recommend the book by Presser et al. as introduction.

References

Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., and Smith, T.W. (2010). Survey Methods in Multinational, Multiregional, and Multicultural Contexts. New York: Wiley.

Hox, J.J., De Leeuw, E.D., and Brinkhuis, M.J.S. (2010). In Survey Methods in Multinational, Multiregional, and Multicultural Contexts, J.A. Harkness, M. Braun, B. Edwards, T.P. Johnson, L. Lyberg, P.P. Mohler, B.-E. Pennell, and T.W. Smith (eds). New York: Wiley.

Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., and Singer, E. (2004). Methods for Testing and Evaluating Survey Questionnaires. New York: Wiley.

Vandenberg, R.J. and Lance, C.E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices and Recommendations for Organizational Research. Organizational Research Methods, 3, 4–69.

Willis, G.B. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage.

Edith de Leeuw
Department of Methodology and Statistics
Faculty of Social and Behavioral Sciences
Utrecht University
PO Box 80.140
NL-3508 TC Utrecht
the Netherlands
Telephone: +31 (0)30 253 4438
E-mail: e.d.deleeuw@uu.nl