# Book Review: Advances in Questionnaire Design, Development, Evaluation and Testing

## Beatty, P. C., Collins, D., Kaye, L., Padilla, J. L., Willis, G. B., & Wilmot, A. (Eds.). (2019). Advances in Questionnaire Design, Development, Evaluation and Testing. John Wiley & Sons. ISBN: 978-1-119-26362-3, December 2019, 816 Pages

**Hong Shi**

China University of Petroleum-Beijing

**Correspondence concerning this article should be addressed to:**

**China University of Petroleum-Beijing**

**E-mail:** hzs0032@tigermail.auburn.edu

# 1 INTRODUCTION

Research on questionnaire design, development, evaluation, and testing has developed significantly over the past decades, triggered by the fact that "survey questionnaires continue to play a unique role in producing data on attitudes and behaviors for which there is no viable alternative"(p. xxv). The changing environment required rethinking the strengths and weaknesses of existing data collection methods. As a result, the chapters have focused on recent changes and challenges and how to enhance the viability of survey questionnaires. Methods are used to complement each other, and qualitative methods are often combined with quantitative methods to develop or assess questionnaires (Persson et al. 2015; Yan et al. 2012). It is in this context that Beatty, Collins, Kaye, Padilla, Willis, and Wilmot's book offers its readers a detailed diachronic view of questionnaire design, development, evaluation, and testing. We are living and working in a more challenging time for questionnaire development and assessment. The approaches, methods, and resources provided in the book will be highly valuable to researchers designing, evaluating, testing, or validating survey questions. Findings and innovations presented from case studies in the book provide useful insight or practicable and effective solutions for survey researchers, projects, organizations, and industries.

   This book is very timely and insightful since the questionnaire issues were explored within a specific context or situation. The authors delineate background on the survey used, describe the key challenges faced in developing a survey, provide a fuller picture of potential data quality problems, and propose guidelines, measures, or recourses in different forms that can be tailored to the "needs" of a particular survey. These suggestions, recommendations, or implications can contribute to an increasing number of studies, whether at the pretesting, main study, or follow-up stage. The potential readership of this book is a wide-ranging one, as it may include survey researchers, practitioners, agencies, and institutes. The book may be particularly helpful to researchers focusing on data quality problems. Detailed methodological accounts of data collection, data analysis procedures, and statistical treatment in database research in different contexts or situations may be especially beneficial for graduate students and early-career researchers. The book also contains numerous and very informative figures and tables that corroborate variation and change of questionnaire design, evaluation, and testing.

   The book is organized into five parts and 30 chapters that address several interrelated aspects of questionnaires. In part I (Assessing the Current Methodology for Questionnaire Design, Development, Testing, and Evaluation), the authors review past, present, and potential future trends influencing questionnaire development, evaluation, and testing (QDET), and propose six needed changes in questionnaire evaluation and testing methods: emphasize survey self-administration as well as interview modes of data collection; evaluate motivational qualities of questionnaires, as well as their cognitive aspects; evaluate the complexities associated with visual communication as well as wording used in oral communication; utilize unified-mode question construction across

survey modes; evaluate whether the equivalent measurement is achieved when mixed-mode surveys are used; design question for use in smartphone and other survey modes. In Chapter 3 the authors give a brief introduction of the methods for questionnaire development, testing, and recommendations, which include expert methods, laboratory methods, field-based methods, usability testing, and statistical modeling techniques. The remaining chapters (Chapters 4 and 5) underline the importance of using various methods including expert review, the Questionnaire Appraisal System (QAS), the Question Understanding Aid (QUAID), the Survey Quality Predictor (SQP) and cognitive interviews to understand the quality of survey questions before fielding them. As the authors and researchers such as Maitland and Presser (2016, 2018) rightly argue, evaluation methods are complementary, and multiple methods can be used instead of relying on a single method. Case studies combined different question testing methods in different ways are presented to show how question-testing aims, choice of question testing methods, and survey life-cycle stage may influence findings.

Part II (Question Characteristics, Response Burden, and Data Quality), including Chapters 6-11, provides an overview of some approaches used to measure, code, and evaluate question characteristics. A taxonomy is highlighted to summarize classes of characteristics that affect the task faced by respondents and interviewers and the liability and validity of data. Readers could find a very detailed description of two case studies in Chapter 6, examining the effects of question characteristics. Chapter 7 builds and expands the multilevel models to understand the joint effects of question and respondent characteristics, and interviewer performance measures on data quality using a large, nationally representative survey. Chapter 8 defines respondent burden as a subjective, multidimensional phenomenon, and underscores that the impact of any given survey feature on burden will vary across respondents based on measurable subjective reactions and attitudes. Structural equation modeling (SEM) is selected as an analytic method to test a model of the burden that includes latent factors related to survey features, respondent perceptions, and other respondent characteristics, and to examine the causal relations (direct and indirect) between these factors and burden. Experiments are presented in Chapter 9 to show the effects of various question characteristics. The authors argue that respondents are sensitive to design characteristics. That design features of survey items can elevate or reduce the motivated underreporting found in earlier studies, in particular when the salience of the filter-follow-up structure and its perceived burden is reduced. The remaining chapters (Chapters 10 and 11) are centered around format issues of questionnaires.

Part III (Chapter 12-18) examines how to improve questionnaires on the web and mobile ⬚devices. With the development of technology and the establishment of internet-based systems and social media applications, new methodological strategies can be used to supplement existing evaluation methods. Chapter 12 introduces emerging survey pretesting methodologies, including usability testing, eye tracking, and online pretesting. It compares these with traditional methods to consider whether these methods can be combined to meet today's web and mobile survey needs. The current practices in usability testing online surveys at the Census Bureau are shared in Chapter 13. The authors highlight both the techniques that offered insights into measuring and improving usability and those that did not appear in this chapter. Areas of agreement between the experience and empirical research evidence are discussed. Readers can find the effects of screen size, design choices, mouse tracking, cognitive probes, and web probing on data quality and how these methods can be best formatted and implemented for mobile device respondents or in web surveys in the remaining chapters (Chapters 14 to 18). The authors conclude that questionnaire evaluation tools such as web probing has strengths and weaknesses, and it most likely "augments", rather than "replaces", traditional cognitive

interviewing.

The Chapters 19-22, included in Part IV (Cross-Cultural and Cross-National Questionnaire Design and Evaluation), all revolve around cross-national/cultural surveys. Concerning cross-national/cultural surveys, their design, development, and evaluation become more complicated and more challenging. Some approaches, practices, guidelines, or resources are recommended by the authors to provide high-quality and comparable measures.

The authors of Chapters 23-30 in Part V (Extensions and Applications) offer some new innovative approaches or directions. All the methods are illustrated with exemplar implementations. For example, in Chapter 23, the authors propose a new approach– regression-based response probing – that combines the strengths of intensive small-sample qualitative approaches with the inferential power of large-scale field trials and experimental manipulations. Chapter 24 opens with a diachronic cover view of contemporary validity theory and validation practices in psychometrics. The authors' highlight contrasts with survey research and methodology, and to bridge psychometrics and survey research; they delineate two approaches to measurement validity: the ecological model of responding to survey questions, and the argument-based approach to validation. Chapter 25 provides a rich account of four large-scale cognitive testing projects, and the quality-driven management approaches are described and explained in detail. The large-scale redesign efforts are illustrated with a research project--the health insurance module in the Current Population Survey (CPS) in Chapter 26. The research project demonstrated that "a small-scale test with a specific objective that is linked to a larger methodological problem can be a much more cost-effective and efficient approach to a major redesign" (p.663). Chapter 27 investigated the gap between time diaries and stylized questions, two common methods of collecting time-use data, and potential sources of measurement error associated with them. A multi-method approach is recommended for future research. Then questionnaire design issues in mail surveys are underlined by the authors in Chapter 28. Chapter 29 goes further to describe the strengths and weaknesses of common testing methods and, more importantly, how and why others can combine them to create a better questionnaire evaluation than any individual test. The reader finds examples in Chapter 29 to illustrate how the bento box approach can be applied to improve questionnaire testing efforts. The authors conclude that for any large and complex data collection, a mixed-methods bento box approach can be more effective than any single method to provide a rich account of potential data quality problems in questionnaire testing. Chapter 30 focuses on questionnaire pretesting from the illustration of a case study. Since it is very complicated, multiple evaluation metrics and evidence sources can be used to assess questions in the pretest.

Case studies presented in the book provide clearer guidance on when which of these methods can or should be used. Considering that the authors' arguments are both theoretically and empirically grounded, these approaches, methods, or resources will be highly insightful and valuable for designing, evaluating, testing, or validating survey questions. The growing practice of utilizing multiple methods illustrates the fact that these different types of approaches are complementary means of investigating the complex phenomena. In the current changing environment, after learning the attractive features and natural limits of existing qualitative and quantitative approaches and the key challenges we faced in developing a survey, we can use multiple methods sequentially, or in parallel, at many decision points to pretest a method, implement it, interpret results, and apply them to refine survey questionnaires.

## 2   References

Maitland, A. & Presser, S. (2016). How accurately do different evaluation methods predict the reliability of survey questions? *Journal of Survey Statistics and Methodology* 4: 362–381.

Maitland, A. & Presser, S. (2018). How do question evaluation methods compare in predicting problems observed in typical survey administration? *Journal of Survey Statistics and Methodology*, 6(4),465-490.

Persson, A., Björnram, A., Elvers, E., & Erikson, J. (2015). A strategy to test questionnaires at a national statistical office.*Statistical Journal of the IAOS*,31(2): 297–304.

Yan, T., Kreuter, F., & Tourangeau, R. (2012). Evaluating survey questions: a comparison of methods. *Journal of Official Statistics*,28: 503–529.

## Acknowledgments

## Funding

## Ethics Declarations

## Competing Interests

No, there are no conflicting interests.

## Rights and Permissions

### Open Access