



SURVEY METHODS IN MULTINATIONAL, MULTIREGIONAL, AND MULTICULTURAL CONTEXTS

Edited by

Janet A. Harkness • Michael Braun • Brad Edwards
Timothy P. Johnson • Lars Lyberg
Peter Ph. Mohler • Beth-Ellen Pennell • Tom W. Smith

 **WILEY**

The Wiley logo consists of a stylized 'W' and 'J' intertwined within a circle, followed by the word 'WILEY' in a bold, serif font.

A JOHN WILEY & SONS, INC., PUBLICATION

This Page Intentionally Left Blank

**SURVEY METHODS
IN MULTINATIONAL,
MULTIREGIONAL, AND
MULTICULTURAL CONTEXTS**

WILEY SERIES IN SURVEY METHODOLOGY

Established in Part by WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors: *Mick P. Couper, Graham Kalton, J. N. K. Rao, Norbert Schwarz,
Christopher Skinner*

Editor Emeritus: *Robert M. Groves*

A complete list of the titles in this series appears at the end of this volume.

SURVEY METHODS IN MULTINATIONAL, MULTIREGIONAL, AND MULTICULTURAL CONTEXTS

Edited by

Janet A. Harkness • Michael Braun • Brad Edwards
Timothy P. Johnson • Lars Lyberg
Peter Ph. Mohler • Beth-Ellen Pennell • Tom W. Smith

 **WILEY**

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2010 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data is available.

ISBN 978-0470-17799-0

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTENTS

CONTRIBUTORS..... ix

PREFACE xiii

PART I SETTING THE STAGE

- 1 **Comparative Survey Methodology**..... 3
Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith
- 2 **Equivalence, Comparability, and Methodological Progress**..... 17
Peter Ph. Mohler and Timothy P. Johnson

PART II QUESTIONNAIRE DEVELOPMENT

- 3 **Designing Questionnaires for Multipopulation Research** 33
Janet A. Harkness, Brad Edwards, Sue Ellen Hansen, Debra R. Miller, and Ana Villar
- 4 **Creating Questions and Protocols for an International Study of Ideas About Development and Family Life** 59
Arland Thornton, Alexandra Achen, Jennifer S. Barber, Georgina Binstock, Wade M. Garrison, Dirgha J. Ghimire, Ronald Inglehart, Rukmalie Jayakody, Yang Jiang, Julie de Jong, Katherine King, Ron J. Lesthaeghe, Sohair Mehanna, Colter Mitchell, Mansoor Moaddel, Mary Beth Ofstedal, Norbert Schwarz, Guangzhou Wang, Yu Xie, Li-Shou Yang, Linda C. Young-DeMarco, and Kathryn Yount
- 5 **Managing the Cognitive Pretesting of Multilingual Survey Instruments: A Case Study of Pretesting of the U.S. Census Bureau Bilingual Spanish/English Questionnaire** 75
Patricia L. Goerman and Rachel A. Caspar
- 6 **Cognitive Interviewing in Non-English Languages: A Cross-Cultural Perspective**..... 91
Yuling Pan, Ashley Landreth, Hyunjoo Park, Marjorie Hinsdale-Shouse, and Alisú Schoua-Glusberg

PART III TRANSLATION, ADAPTATION, AND ASSESSMENT

- 7 **Translation, Adaptation, and Design** 117
Janet A. Harkness, Ana Villar, and Brad Edwards
- 8 **Evaluation of a Multistep Survey Translation Process** 141
Gordon B. Willis, Martha Stapleton Kudela, Kerry Levin, Alicia Norberg, Debra S. Stark, Barbara H. Forsyth, Pat Dean Brick, David Berrigan, Frances E. Thompson, Deirdre Lawrence, and Anne M. Hartman

9 Developments in Translation Verification Procedures in Three Multilingual Assessments: A Plea for an Integrated Translation and Adaptation Monitoring Tool..... 157
Steve Dept, Andrea Ferrari, and Laura Wäyrynen

PART IV CULTURE, COGNITION, AND RESPONSE

10 Cognition, Communication, and Culture: Implications for the Survey Response Process 177
Norbert Schwarz, Daphna Oyserman, and Emilia Peytcheva

11 Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey-Response Process..... 191
Ayşe K. Uskul, Daphna Oyserman, and Norbert Schwarz

12 Response Styles and Culture..... 203
Yongwei Yang, Janet A. Harkness, Tzu-Yun Chin, and Ana Villar

PART V KEY PROCESS COMPONENTS AND QUALITY

13 Quality Assurance and Quality Control in Cross-National Comparative Studies..... 227
Lars Lyberg and Diana Maria Stukel

14 Sampling Designs for Cross-Cultural and Cross-National Survey Programs 251
Steven G. Heeringa and Colm O’Muircheartaigh

15 Challenges in Cross-National Data Collection 269
Beth-Ellen Pennell, Janet A. Harkness, Rachel Levenstein, and Martine Quaglia

16 A Survey Process Quality Perspective on Documentation 299
Peter Ph. Mohler, Sue Ellen Hansen, Beth-Ellen Pennell, Wendy Thomas, Joachim Wackerow, and Frost Hubbard

17 Harmonizing Survey Data 315
Peter Granda, Christof Wolf, and Reto Hadorn

PART VI NONRESPONSE

18 The Use of Contact Data in Understanding Cross-National Differences in Unit Nonresponse..... 335
Annelies G. Blom, Annette Jäckle, and Peter Lynn

19 Item Nonresponse and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective 355
Joachim R. Frick and Markus M. Grabka

PART VII ANALYZING DATA

20 An Illustrative Review of Techniques for Detecting Inequivalences..... 375
Michael Braun and Timothy P. Johnson

21 Analysis Models for Comparative Surveys..... 395
Joop J. Hox, Edith de Leeuw, and Matthieu J.S. Brinkhuis

22 Using Polytomous Item Response Theory to Examine Differential Item and Test Functioning: The Case of Work Ethic 419
David J. Woehr and John P. Meriac

23 Categorization Errors and Differences in the Quality of Questions in Comparative Surveys 435
Daniel Oberski, Willem E. Saris, and Jacques A. Hagenaars

24 Making Methods Meet: Mixed Design in Cross-Cultural Research 455
Fons J.R. van de Vijver and Athanasios Chasiotis

PART VIII GLOBAL SURVEY PROGRAMS

25 The Globalization of Survey Research..... 477
Tom W. Smith

26 Measurement Equivalence in Comparative Surveys: the European Social Survey (ESS)—From Design to Implementation and Beyond..... 485
Rory Fitzgerald and Roger Jowell

27 The International Social Survey Programme: Annual Cross-National Surveys Since 1985..... 497
Knut Kalgraff Skjåk

28 Longitudinal Data Collection in Continental Europe: Experiences from the Survey of Health, Ageing, and Retirement in Europe (SHARE) 507
Axel Börsch-Supan, Karsten Hank, Hendrik Jürges, and Mathis Schröder

29 Assessment Methods in IEA’s TIMSS and PIRLS International Assessments of Mathematics, Science, and Reading..... 515
Ina V. S. Mullis and Michael O. Martin

30 Enhancing Quality and Comparability in the Comparative Study of Electoral Systems (CSES) 525
David A. Howell

31 The Gallup World Poll..... 535
Robert D. Tortora, Rajesh Srinivasan, and Neli Esipova

REFERENCES 545

SUBJECT INDEX 595

This Page Intentionally Left Blank

CONTRIBUTORS

Alexandra Achen, University of Michigan, Ann Arbor, Michigan
Jennifer S. Barber, University of Michigan, Ann Arbor, Michigan
David Berrigan, National Cancer Institute, Bethesda, Maryland
Georgina Binstock, Cenep-Conicet, Argentina
Annelies G. Blom, Mannheim Research Institute for the Economics of Aging,
Mannheim, Germany
Axel Börsch-Supan, Mannheim Research Institute for the Economics of Aging,
Mannheim, Germany
Michael Braun, GESIS–Leibniz Institute for the Social Sciences, Mannheim,
Germany
Pat Dean Brick, Westat, Rockville, Maryland
Matthieu J. S. Brinkhuis, Utrecht University, Utrecht, The Netherlands
Rachel A. Caspar, RTI, Research Triangle Park, North Carolina
Athanasios Chasiotis, Tilburg University, Tilburg, The Netherlands
Tzu-Yun Chin, University of Nebraska-Lincoln, Lincoln, Nebraska
Julie de Jong, University of Michigan, Ann Arbor, Michigan
Edith de Leeuw, Utrecht University, Utrecht, The Netherlands
Steve Dept, cApStAn Linguistic Quality Control, Brussels, Belgium
Brad Edwards, Westat, Rockville, Maryland
Neli Esipova, Gallup, Inc., Princeton, New Jersey
Andrea Ferrari, cApStAn Linguistic Quality Control, Brussels, Belgium
Rory Fitzgerald, City University, London, United Kingdom
Barbara H. Forsyth, University of Maryland Center for the Advanced Study of
Language, Maryland
Joachim R. Frick, DIW Berlin, University of Technology Berlin, IZA Bonn,
Germany
Wade M. Garrison, University of Kansas, Lawrence, Kansas
Dirgha J. Ghimire, University of Michigan, Ann Arbor, Michigan
Patricia L. Goerman, U.S. Bureau of the Census, Washington, DC
Markus Grabka, DIW Berlin, University of Technology Berlin, Germany
Peter Granda, ICPSR, University of Michigan, Ann Arbor, Michigan
Reto Hadorn, SIDOS, Neuchatel, Switzerland
Jacques A. Hagenaars, Tilburg University, Tilburg, The Netherlands
Karsten Hank, Mannheim Research Institute for the Economics of Aging,
Mannheim, Germany
Sue Ellen Hansen, University of Michigan, Ann Arbor, Michigan
Janet A. Harkness, University of Nebraska-Lincoln, Lincoln, Nebraska
Anne M. Hartman, National Cancer Institute, Bethesda, Maryland
Steven G. Heeringa, University of Michigan, Ann Arbor, Michigan
Marjorie Hinsdale-Shouse, RTI, Research Triangle Park, North Carolina
David A. Howell, University of Michigan, Ann Arbor, Michigan
Joop J. Hox, Utrecht University, Utrecht, The Netherlands
Frost Hubbard, University of Michigan, Ann Arbor, Michigan
Ronald Inglehart, University of Michigan, Ann Arbor, Michigan

- Annette Jäckle**, Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom
- Rukmalie Jayakody**, Pennsylvania State University, University Park, Pennsylvania
- Yang Jiang**, University of Michigan, Ann Arbor, Michigan
- Timothy P. Johnson**, University of Illinois at Chicago, Chicago, Illinois
- Roger Jowell**, City University, London, United Kingdom
- Hendrik Jürges**, Mannheim Research Institute for the Economics of Aging, Mannheim, Germany
- Knut Kalgraff Skjåk**, Norwegian Social Science Data Services, Bergen, Norway
- Katherine King**, University of Michigan, Ann Arbor, Michigan
- Ashley Landreth**, U.S. Bureau of Census, Washington, DC
- Deirdre Lawrence**, National Cancer Institute, Bethesda, Maryland
- Ron J. Lesthaeghe**, Belgian Royal Academy of Sciences, Brussels, Belgium
- Rachel Levenstein**, University of Michigan, Ann Arbor, Michigan
- Kerry Levin**, Westat, Rockville, Maryland
- Lars Lyberg**, Statistics Sweden and Stockholm University, Stockholm, Sweden
- Peter Lynn**, Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom
- Michael O. Martin**, Lynch School of Education, Boston College, Chestnut Hill, Massachusetts
- Sohair Mehanna**, American University in Cairo, Cairo, Egypt
- John P. Meriac**, University of Missouri-St. Louis, St. Louis, Missouri
- Debra R. Miller**, University of Nebraska-Lincoln, Lincoln, Nebraska
- Colter Mitchell**, Princeton University, Princeton, New Jersey
- Mansoor Moaddel**, Eastern Michigan University, Ypsilanti, Michigan
- Peter Ph. Mohler**, University of Mannheim, Mannheim, Germany
- Ina V. S. Mullis**, Lynch School of Education, Boston College, Chestnut Hill, Massachusetts
- Alicia Norberg**, Westat, Rockville, Maryland
- Colm O'Muircheartaigh**, National Opinion Research Center and University of Chicago, Chicago, Illinois
- Daniel Oberski**, Tilburg University, Tilburg, The Netherlands and Universitat Pompeu Fabra, Barcelona, Spain
- Mary Beth Ofstedal**, University of Michigan, Ann Arbor, Michigan
- Daphna Oyserman**, University of Michigan, Ann Arbor, Michigan
- Yuling Pan**, U.S. Bureau of Census, Washington, DC
- Hyunjoo Park**, RTI, Research Triangle Park, North Carolina
- Beth-Ellen Pennell**, University of Michigan, Ann Arbor, Michigan
- Emilia Peytcheva**, RTI, Research Triangle Park, North Carolina
- Martine Quaglia**, Institut National d'Etudes Démographiques, Paris, France
- Willem E. Saris**, ESADE, Universitat Pompeu Fabra, Barcelona, Spain
- Alisú Schoua-Glusberg**, Research Support Services, Evanston, Illinois
- Mathis Schröder**, Mannheim Research Institute for the Economics of Aging, Mannheim, Germany
- Norbert Schwarz**, University of Michigan, Ann Arbor, Michigan

- Tom W. Smith**, National Opinion Research Center, University of Chicago, Chicago, Illinois
- Rajesh Srinivasan**, Gallup, Inc., Princeton, New Jersey
- Martha Stapleton Kudela**, Westat, Rockville, Maryland
- Debra S. Stark**, Westat, Rockville, Maryland
- Diana Maria Stukel**, Westat, Rockville, Maryland
- Wendy Thomas**, University of Minnesota, Minneapolis, Minnesota
- Frances E. Thompson**, National Cancer Institute, Bethesda, Maryland
- Arland Thornton**, University of Michigan, Ann Arbor, Michigan
- Robert D. Tortora**, Gallup, Inc., London, United Kingdom
- Ayşe K. Uskul**, Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom
- Fons J.R. van de Vijver**, Tilburg University, Tilburg, The Netherlands, and North-West University, South Africa
- Ana Villar**, University of Nebraska-Lincoln, Lincoln, Nebraska
- Joachim Wackerow**, GESIS–Leibniz Institute for the Social Sciences, Mannheim, Germany
- Guangzhou Wang**, Chinese Academy of Social Sciences, Beijing, China
- Laura Wäyrynen**, cApStAn Linguistic Quality Control, Brussels, Belgium
- Gordon Willis**, National Cancer Institute, Bethesda, Maryland
- David J. Woehr**, University of Tennessee-Knoxville, Knoxville, Tennessee
- Christof Wolf**, GESIS–Leibniz Institute for the Social Sciences, Mannheim, Germany
- Yu Xie**, University of Michigan, Ann Arbor, Michigan
- Li-Shou Yang**, University of Michigan, Ann Arbor, Michigan
- Yongwei Yang**, Gallup, Inc., Omaha, Nebraska
- Linda C. Young-DeMarco**, University of Michigan, Ann Arbor, Michigan
- Kathryn Yount**, Emory University, Atlanta, Georgia

This Page Intentionally Left Blank

PREFACE

This book aims to provide up-to-date insight into key aspects of methodological research for comparative surveys. In conveying information about cross-national and cross-cultural research methods, we have often had to assume our readers are comfortable with the essentials of basic survey methodology. Most chapters emphasize multinational research projects. We hope that in dealing with more complex and larger studies, we address many of the needs of researchers engaged in within-country research.

The book both precedes and follows a conference on Multinational, Multicultural, and Multiregional Survey Methods (3MC) held at the Berlin Brandenburg Academy of Sciences and Humanities in Berlin, June 25–28, 2008. The conference was the first international conference to focus explicitly on survey methodology for comparative research (<http://www.3mc2008.de/>). With the conference and monograph, we seek to draw attention to important recent changes in the comparative methodology landscape, to identify new methodological research and to help point the way forward in areas where research needs identified in earlier literature have still not been adequately addressed.

The members of the Editorial Committee for the book, chaired by Janet Harkness, were: Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith. Fons van de Vijver joined this team to complete the conference Organizing Committee. The German Federal Minister of Education and Research, Frau Dr. Annette Schavan, was patron of the conference, and we were fortunate to have Denise Lievesley, Lars Lyberg, and Sidney Verba as keynote speakers for the opening session in the splendid Konzerthaus Berlin.

The conference and book would not have been possible without funding from sponsors and donors we acknowledge here and on the conference website.

Five organizations sponsored the conference and the preparation of the monograph. They are, in alphabetical order: the American Association for Public Opinion Research; the American Statistical Association (Survey Research Methods Section); Deutsche Forschungsgemeinschaft [German Science Foundation]; what was then Gesis-ZUMA (The Centre for Survey Research and Methodology, Mannheim); and Survey Research Operations, Institute for Social Research, at the University of Michigan. These organizations variously provided funds, personnel, and services to support the development of the conference and the monograph.

Additional financial support for the conference and book was provided by, in alphabetical order:

Eurostat
GfK Group
Interuniversity Consortium for Political and Social Research (ICPSR)
(U.S.) National Agricultural Statistics Services

Radio Regenbogen Hörfunk in Baden
The Roper Center
SAP
Statistics Sweden
TNS Political & Social
U.S. Census Bureau

Nielsen
NORC at the University of Chicago
PEW Research Center

University of Nebraska at Lincoln
Westat

We must also thank the Weierstraß Institute, Berlin, for donating the use of its lecture room for conference sessions, the Studienstiftung des Deutschen Volkes [German National Academic Foundation] for providing a room within the Berlin Brandenburg Academy for the conference logistics, and colleagues from the then Gesis-Aussenstelle, Berlin, for their support at the conference.

Planning for the book and the conference began in 2006. Calls for papers for consideration for the monograph were circulated early in 2007. A large number of proposals were received; some 60% of the book is derived from these. At the same time, submissions in some topic areas were sparse. We subsequently sought contributions for these underrepresented areas, also drawing from the ranks of the editors. In this way, the book came to consist both of stand-alone chapters which aim to treat important topics in a global fashion and sets of chapters that illustrate different facets of a topic area. The result, we believe, reflects current developments in multinational, multilingual, and multicultural research.

The book is divided into eight parts:

- I Setting the Stage
- II Questionnaire Development
- III Translation, Adaptation, and Assessment
- IV Culture, Cognition, and Response
- V Key Process Components and Quality
- VI Nonresponse
- VII Analyzing Data
- VIII Global Survey Programs

Chapter 1 looks at comparative survey methodology for today and tomorrow, and Chapter 2 discusses fundamental aspects of this comparative methodology. Chapters 3 and 4 consider question and questionnaire design, Chapter 3 from a global perspective and Chapter 4 from a study-specific standpoint. Chapters 5 and 6 discuss developments in pretesting translated materials. Chapter 5 moves toward some guidelines on the basis of lessons learned, and Chapter 6 applies discourse analysis techniques in pretesting. Chapters 7, 8, and 9 focus on producing and evaluating survey translations and instrument adaptations. Chapter 7 is on survey translation and adaptation; Chapter 8 presents a multistep procedure for survey translation assessment; and Chapter 9 describes translation verification strategies for international educational testing instruments. Chapters 10 and 11 consider cultural differences in how information in questions and in response categories is perceived and processed and the relevance for survey response. Together with Chapter 12, on response styles in cultural contexts, they complement and expand on points made in earlier parts.

Part V presents a series of chapters that deal with cornerstone components of the survey process. Chapter 13 outlines the quality framework needed for

multipopulation research; Chapter 14 discusses cross-national sampling in terms of design and implementation; Chapter 15 is a comprehensive overview of data collection challenges; and Chapter 16 discusses the role of documentation in multipopulation surveys and emerging documentation standards and tools for such surveys. Chapter 17 treats input and output variable harmonization. Each of these chapters emphasizes issues of survey quality from their particular perspective. Part VI consists of two chapters on nonresponse in comparative contexts: Chapter 18 is on unit nonresponse in cross-national research and Chapter 19 is on item nonresponse in longitudinal panel studies. Both these contribute further to the discussion of comparative data quality emphasized in contributions in Part V.

Chapter 20 introduces Part VII, which contains five chapters on analysis in comparative contexts. Chapter 20 demonstrates the potential of various techniques by applying them to a single multipopulation dataset; Chapter 21 treats multigroup and multilevel structural equation modeling and multilevel latent class analysis; Chapter 22 discusses polytomous item response theory; Chapter 23 explores categorization problems and a Multitrait Multimethods (MTMM) design, and the last chapter in the section, Chapter 24, discusses mixed methods designs that combine quantitative and qualitative components.

Part VIII is on global survey research and programs. It opens in Chapter 25 with an overview of developments in global survey research. Chapters 26–31 present profiles and achievements in a variety of global research programs. Chapter 26 is on the European Social Survey; Chapter 27 presents the International Social Survey Programme; Chapter 28 deals with the Survey of Health, Ageing, and Retirement in Europe. Chapter 29 discusses developments in two international education assessment studies, the Trends in International Mathematics and Science Study and the Progress in International Reading Literacy Study. Chapter 30 profiles the Comparative Study of Electoral Systems, and the concluding chapter in the volume, Chapter 31, describes the Gallup World Poll.

Pairs of editors or individual editors served as the primary editors for invited chapters: Edwards and Harkness for Chapters 4, 5, and 6; Harkness and Edwards for Chapters 8 and 9; Braun and Harkness for Chapters 10 and 11; Johnson for Chapter 12; Lyberg for Chapters 14, 18, and 19; Lyberg, Pennell, and Harkness for Chapter 17; Braun and Johnson for Chapters 21–24; Smith for Chapters 26–31. The editorial team also served as the primary reviewers of chapters in which editors are only or first author (Chapters 2, 3, 7, 13, 15, 16, 20, and 25).

The editors have many people to thank and acknowledge. First, we thank our authors for their contributions, their perseverance, and their patience. We also thank those who helped produce the technical aspects of the book, in particular Gail Arnold at ISR, University of Michigan, who formatted the book, Linda Beatty at Westat who designed the cover, Peter Mohler, who produced the subject index, and An Lui, Mathew Stange, Clarissa Steele, and, last but not least, Ana Villar, all at the University of Nebraska-Lincoln, who aided Harkness in the last phases of completing the volume. In addition, we would like to thank Fons van de Vijver, University of Tilburg, Netherlands, for reviewing Chapter 12 and students from Harkness' UNL spring 2009 course for their input on Chapters 13 and 15. We are grateful to our home organizations and institutions for enabling us to work on the volume and, as relevant, to host or attend editorial meetings. Finally, we thank

Steven Quigley and Jacqueline Palmieri at Wiley for their support throughout the production process and their always prompt attendance to our numerous requests.

Janet A. Harkness
Michael Braun
Timothy P. Johnson
Lars Lyberg
Peter Ph. Mohler
Beth-Ellen Pennell
Tom W. Smith

PART I

SETTING THE STAGE

This Page Intentionally Left Blank

1

Comparative Survey Methodology

*Janet A. Harkness, Michael Braun, Brad Edwards,
Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler,
Beth-Ellen Pennell, and Tom W. Smith*

1.1 INTRODUCTION

This volume discusses methodological considerations for surveys that are deliberately designed for comparative research such as multinational surveys. As explained below, such surveys set out to develop instruments and possibly a number of the other components of the study specifically in order to collect data and compare findings from two or more populations.

As a number of chapters in this volume demonstrate, multinational survey research is typically (though not always) more complex and more complicated to undertake successfully than are within-country cross-cultural surveys. Many chapters focus on this more complicated case, discussing multinational projects such as the annual International Social Survey Programme (ISSP), the epidemiologic World Mental Health Initiative survey (WMH), the 41-country World Fertility Survey (WFS), or the triennial and worldwide scholastic assessment Programme for International Student Assessment (PISA). Examples of challenges and solutions presented in the volume are often drawn from such large projects.

At the same time, we expect many of the methodological features discussed here also to apply for within-country comparative research as well. Thus we envisage chapters discussing question design, pretesting, translation, adaptation, data collection, documentation, harmonization, quality frameworks, and analysis to provide much of importance for within-country comparative researchers as well as for those involved in cross-national studies.

This introductory chapter is organized as follows. Section 1.2 briefly treats the growth and standing of comparative surveys. Section 1.3 indicates overlaps between multinational, multilingual, multicultural, and multiregional survey

¹ Survey Methods in Multinational, Multiregional, and Multicultural Contexts, edited by Harkness et al. Copyright © 2010 John Wiley & Sons, Inc.

research and distinguishes between comparative research and surveys deliberately designed for comparative purposes. Section 1.4 considers the special nature of comparative surveys, and Section 1.5 how comparability may drive design decisions. Section 1.6 considers recent changes in comparative survey research methods and practice. The final section, 1.7, considers ongoing challenges and the current outlook.

1.2 COMPARATIVE SURVEY RESEARCH: GROWTH AND STANDING

Almost without exception, those writing about comparative survey research—whether from the perspective of marketing, the social, economic and behavioral sciences, policy-making, educational testing, or health research—remark upon its “rapid,” “ongoing,” or “burgeoning” growth. And in each decade since World War II, a marked “wave” of interest in conducting cross-national and cross-cultural survey research can be noted in one discipline or another (see contributions in Bulmer, 1998; Bulmer & Warwick, 1983/1993; Gauthier, 2002; Hantrais, 2009; Hantrais & Mangen, 2007; Øyen, 1990; and Chapters 2 and 25, this volume).

Within the short span of some 50 years, multipopulation survey research has become accepted as not only useful and desirable but, indeed, as indispensable. In as much as international institutions and organizations—such as the European Commission, the Organization for Economic Co-operation and Development (OECD), the United Nations (UN), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the World Bank, the International Monetary Fund (IMF), and the World Health Organization (WHO)—depend on multinational data to inform numerous activities, it has become ubiquitous and, in some senses, also commonplace.

1.3 TERMINOLOGY AND TYPES OF RESEARCH

In this section we make a distinction which is useful for the special methodological focus of many chapters in this volume—between comparative research in general and deliberately designed comparative surveys.

1.3.1 Multipopulation Surveys: Multilingual, Multicultural, Multinational, and Multiregional

Multipopulation studies can be conducted in one language; but most multipopulation research is nonetheless also multilingual. At the same time, cultural differences exist between groups that share a first language both within a country (e.g., the Welsh, Scots, Northern Irish, and English in the United Kingdom) and across countries (e.g., French-speaking nations/populations). Language difference (Czech versus Slovakian, Russian versus Ukrainian) is, therefore, not a necessary prerequisite for cultural difference, but it is a likely indicator of cultural difference.

Within-country research can be multilingual, as reflected in national research conducted in countries as different as the Philippines, the United States, Switzerland, Nigeria, or in French-speaking countries in Africa. Cross-national projects may thus often need to address within-country differences in language and culture in addition to across-country differences, both with respect to instrument versions and norms of communication.

Multiregional research may be either within- or across-country research and the term is used flexibly. Cross-national multiregional research may group countries considered to “belong together” in some respect, such as geographical location (the countries of Meso and Latin America), in demographical features (high or low birth or death rates, rural or urban populations), in terms of developmental theory (see Chapter 4, this volume) or in terms of income variability. Other multiregional research might be intent on covering a variety of specific populations in different locations or on ensuring application in a multitude of regions and countries. Within-country multiregional research might compare differences among populations in terms of north-south, east-west or urban-rural divisions.

1.3.2 Comparative by Design

This volume focuses on methodological considerations for surveys that are deliberately planned for comparative research. These are to be understood as projects that deliberately design their instruments and possibly other components of the survey in order to compare different populations and that collect data from two or more different populations. In 1969, Stein Rokkan commented on the rarity of “deliberately designed cross-national surveys” (p. 20). Comparative survey research has grown tremendously over the last four decades and is ubiquitous rather than rare. However, Rokkan’s warning that these surveys are not “surefire investments” still holds true; the success of any comparative survey requires to be demonstrated and cannot be assumed simply on the basis of protocols or specifications followed. Numerous chapters in this volume address how best to construct and assess different aspects of surveys designed for comparative research.

Comparative instruments are manifold in format and purpose: educational or psychological tests, diagnostic instruments for health, sports performance, needs or usability assessment tools; social science attitudinal, opinion and behavioral questionnaires; and market research instruments to investigate preferences in such things as size, shape, color, or texture. Several chapters also present comparative methodological studies.

Comparative surveys are conducted in a wide variety of modes, can be longitudinal, can compare different populations across countries or within countries, and can be any mix of these. Some of the studies referred to in the volume are longitudinal in terms of populations studied (panels) or in terms of the contents of the research project (programs of replication). Most of the methodological discussion here, however, focuses on synchronic, across-population research rather than on across-time perspectives (but see Lynn, 2009; Duncan, Kalton, Kasprzyk, & Singh, 1989; Smith, 2005).

Comparative surveys may differ considerably in the extent to which the deliberate design includes such aspects as sampling, the data collection process, documentation, or harmonization. In some cases, the instrument is the main component “designed” to result in comparable data, while many other aspects are decided at the local level (e.g., mode, sample design, interviewer assignment, and contact protocols). Even when much is decided at the local level, those involved in the project must implicitly consider these decisions compatible with the comparative goals of the study.

If we examine a range of large-scale cross-national studies conducted in the last few decades (see, for example, Chapters 25–31, this volume), marked differences can also be found in study design and implementation. Studies vary greatly in the level of coordination and standardization across the phases of the survey life cycle, for example, in their transparency and documentation of methods, and in their data collection requirements and approaches.

1.3.3 Comparative Uses of National Data

Comparative research (of populations and locations) need not be based on data derived from surveys deliberately designed for that purpose.

A large body of comparative research in official statistics, for instance, is carried out using data from national studies designed for domestic purposes which are then also used in analyses across samples/populations/countries. Early cross-national social science research often consisted of such comparisons (cf. Gauthier, 2002; Mohler & Johnson, this volume; Rokkan, 1969; Scheuch, 1973; Verba, 1969). Official statistics agencies working at national and international levels (UNESCO Statistics; the European statistical agency, Eurostat; and national statistical agencies such as the German Statistisches Bundesamt and Statistics Canada) often utilize such national data for comparative purposes, as do agencies producing international data on labor force statistics (International Labour Organization; ILO), on income, wealth, and poverty (Luxembourg Income Study; LIS), and on employment status (Luxembourg Employment Study; LES). Such agencies harmonize data from national studies and other sources because adequately rich and reliable data from surveys that were deliberately designed to produce cross-national datasets are not available for many topics. The harmonization strategies used to render outputs from national data comparable (ex-post output harmonization) are deliberately designed for that purpose (see, for example, Ehling, 2003); it is the national surveys themselves which are not comparative by design. A partnership between Eurostat and many national statistical offices has resulted in the European Statistical System, an initiative which aims to provide reliable and comparable statistics for all the European Union and the European Free Trade Association Member States on the basis of national data.

Instruments designed for a given population are also frequently translated and fielded with other populations. Such translated versions can be tested for suitability with the populations requiring the translations (see Chapters 5, 6, and 7, this volume) and may produce data that permit comparison. Nonetheless, the

original (source) instrument was not comparative by design. Publications arguing the validity and reliability of “translated/adapted” instruments abound, particularly in health, opinion, and psychological research. While the suitability of these procedures and instruments is sometimes contested (e.g., Greenfield, 1997), such instruments may be translated into many languages and used extensively worldwide. In some cases, feedback from implementations in other languages can lead to adjustments to the original instrument. One prominent example is the development of the SF-36 Health Survey, a short (36-question) survey that has been translated and adapted in over 50 languages. The development of translated versions led to related modifications in the original English questionnaire (cf. Ware, undated, at <http://www.sf-36.org/tools/SF36.shtml/>).

Finally, we note that the distinction between comparative research and research that is comparative by design used here is not one always made. Lynn, Japac, and Lyberg (2006), for example, use the term “cross-national surveys” to refer to “all types of surveys where efforts are made to achieve comparability across countries. Efforts to achieve comparability vary on a wide spectrum from opportunistic adjustment of data after they have been collected to deliberate design of each step in the survey process to achieve functional equivalence” (p. 7). The latter of these would fall under our definition of “surveys comparative by design;” those based on “opportunistic adjustment of data after they have been collected” would not.

1.4 WHAT IS (SO) SPECIAL ABOUT COMPARATIVE SURVEY RESEARCH?

Many discussions of comparative survey research note at some point that all social science research is comparative (cf. Armer, 1973; Jowell, 1998; Lipset, 1986; Smith, forthcoming).

Some also suggest that there is nothing really special about comparative (survey) research. Verba (1971 and 1969) and Armer (1973) seem to take this position—but simultaneously also document difference. Verba (1969) states, for example: “The problems of design for within-nation studies apply for across-nation studies. If the above sentence seems to say there is nothing unique about cross-cultural studies, it is intended. The difference is that the problems are more severe and more easily recognizable” (p. 313). Armer (1973) goes a step further: “My argument is that while the problems involved are no different in kind from those involved in domestic research, they are of such great magnitude as to constitute an almost qualitative difference for comparative as compared to noncomparative research” (p. 4).

Later researchers, focusing more on the design and organization of comparative surveys, point to what they consider to be unique aspects. Lynn, Japac, and Lyberg (2006) suggest “Cross-national surveys can be considered to have an extra layer of survey design, in addition to the aspects that must be considered for any survey carried out in a single country” (p. 17). Harkness, Mohler, and van de Vijver (2003) suggest that different kinds of surveys call for different tools and strategies. Certain design strategies, such as decentering,

certainly have their origin and purpose in the context of developing comparative instruments (cf. Werner & Campbell, 1970). The distinction between comparative research and surveys that are comparative by design accommodates the view that all social science research is comparative and that national data can be used in comparative research, while also allowing for the need for special strategies and procedures in designing and implementing surveys directly intended for comparative research.

There is considerable consensus that multinational research is valuable and also more complex than single-country research (Kohn, 1987; Jowell, 1998; Kuechler, 1998; Lynn, Japiec, & Lyberg, 2006; Rokkan, 1969). The special difficulties often emphasized include challenges to “equivalence,” multiple language and meaning difficulties, conceptual and indicator issues, obtaining good sample frames, practical problems in data collection, as well as the sheer expense and effort involved. A number of authors in the present volume also point to the organizational demands as well as challenges faced in dealing with the varying levels of expertise and the different *modi operandi*, standards, and perceptions likely to be encountered in different locations.

1.5 HOW COMPARABILITY MAY DRIVE DESIGN

The comparative goals of a study may call for special design, process, and tool requirements not needed in other research. Examples are such unique requirements as decentering, or *ex ante* input harmonization (cf. Ehling, 2003). But deliberately designed comparative surveys may also simply bring to the foreground concerns and procedures that are not a prime focus of attention in assumed single-population studies (communication channels, shared understanding of meaning, complex organizational issues, researcher expertise, training, and documentation).

What Lynn and colleagues (2006) conceptualize as a layer can usefully be seen as a central motivation for design and procedures followed, a research and output objective at the hub of the survey life cycle that shapes decisions about any number of the components and procedures of a survey from its organizational structure, funding, working language(s), researcher training, and quality frameworks to instrument design, sample design, data collection modes, data processing, analysis, documentation, and data dissemination. Figure 1.1 is a simplified representation of this notion of comparability *driving* design decisions. For improved legibility, we display only four major components in the circle quadrants, instead of all the life-cycle stages actually involved. The comment boxes outside also indicate only a very few examples of the many trade-offs and other decisions to be made for each component in a comparative-by-design survey.

1.6 RECENT CHANGES IN PRACTICE, PRINCIPLES, AND PERSPECTIVES

The practices followed and tools employed in the design, implementation, and analysis of general (noncomparative) survey-based research have evolved rapidly

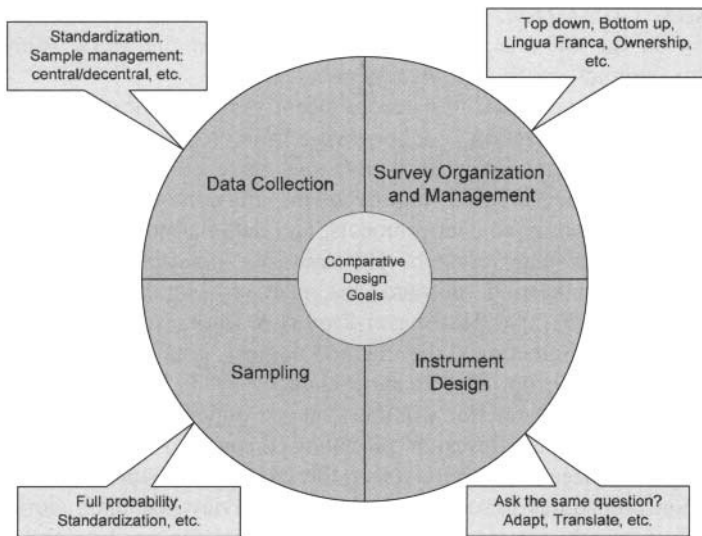


Figure 1.1. “Comparative-by-Design” Surveys

in recent decades. Albeit with some delay, these developments in general survey research methodology are carrying over into comparative survey research. A field of “survey methodology” has emerged, with standardized definitions and (dynamic) benchmarks of good and best practice (cf. Groves et al., 2009, pp. 1–37). Techniques and strategies emerging in the field have altered the way survey research is conceptualized, undertaken, and (now) taught at Masters and PhD level, quietly putting the lie to Scheuch’s (1989) claim (for comparative surveys) that “in terms of methodology *in abstracto* and on issues of research technology, most of all that needed to be said has already been published” (p. 147).

The size and complexity of cross-national and cross-cultural survey research have themselves changed noticeably in the last 15–20 years, as have perspectives on practices and expectations for quality. Large-scale comparative research has become a basic source of information for governments, international organizations, and individual researchers. As those involved in research and analysis have amassed experience, the field has become increasingly self-reflective of procedures, products, and assumptions about good practice. In keeping with this, a number of recent publications discuss the implementation of specific projects across countries. These include Börsch-Supan, Jürges, and Lipps (2003) on the Survey on Health, Ageing and Retirement in Europe; Brancato (2006) on the European Statistical System; Jowell, Roberts, Fitzgerald, and Eva (2007b) on the European Social Survey (ESS); and Kessler and Üstün (2008) on the World Mental Health Initiative (WMH). Manuals and technical reports are available on the implementation of specific studies. Examples are Barth, Gonzalez, and

Neuschmidt (2004) on the Trends in International Mathematics and Science Study (TIMSS); Fisher, Gershuny, and Gauthier (2009) on the Multinational Time Use Study; Grosh and Muñoz (1996) on the Living Standards Measurement Study Survey; IDASA, CDD-Ghana, and IREEP (2007) on the Afrobarometer; ORC Macro (2005) on the Malaria Indicator Survey; and, on the PISA study, the Programme for International Student Assessment (2005).

Comparative methodological research in recent years has turned to questions of implementation, harmonization and, borrowing from cross-cultural psychology, examination of bias. Methodological innovations have come from within the comparative field; evidence-based improvements in cross-cultural pretesting, survey translation, sampling, contact protocols, and data harmonization are a few examples. Recent pretesting research addresses not just the need for *more* pretesting, but for pretesting tailored to meet cross-cultural needs (see contributions in Harkness, 2006; Hoffmeyer-Zlotnik & Harkness, 2005; Chapters 5 and 6, this volume). Sometimes the methodological issues have long been recognized—Verba (1969, pp. 80–99) and Scheuch (1993, 1968, pp. 110, 119) could hardly have been clearer, for example, on the importance of context—but only now is methodological research providing theoretical insights into how culture and context affect perception (see, for example, Chapters 10–12, this volume). Design procedures have come under some review; scholars working in Quality of Life research, for instance, have emphasized the need to orchestrate cross-cultural involvement in instrument design (Fox-Rushby & Parker, 1995; Skevington, 2002).

The increased attention paid to quality frameworks in official statistics comprising, among others, dimensions such as relevance, timeliness, accuracy, comparability, and coherence (Biemer & Lyberg, 2003; Chapter 13, this volume), combined with the “total survey error” (TSE) paradigm (Groves, 1989) in survey research, is clearly carrying over into comparative survey research, despite the challenges this involves (cf. Chapter 13, this volume). Obviously, the comparability dimension has a different meaning in a 3M context than in a national survey and could replace the TSE paradigm as the main planning criterion in such a context, as Figure 1.1 also suggests.

Jowell (1998) remarked on quality discrepancies between standards maintained in what he called “national” research and the practices and standards followed in cross-national research. Jowell’s comments coincided with new initiatives in the International Social Survey Programme to monitor study quality and comparability (cf. Park & Jowell, 1997a) as well as the beginning of a series of publications on comparative survey methods (e.g., Harkness, 1998; Saris & Kaase, 1997) and the development of a European Science Foundation blueprint (ESF, 1999) for the European Social Survey (ESS). The ISSP and the ESS have incorporated study monitoring and methodological research in their programs; both of these ongoing surveys have also contributed to the emergence of a body of researchers whose work often concentrates on comparative survey methods.

Particular attention has been directed recently to compiling guidelines and evidence-based benchmarks, developing standardization schemes, and establishing specifications and tools for quality assurance and quality control in comparative survey research. The cross-cultural survey guidelines at <http://www.ccsq.isr>.

umich.edu/ are a prominent example. Numerous chapters in the volume treat such developments from the perspective of their given topics.

Initiatives to improve comparability and ensure quality are found in other disciplines too. The International Test Commission has, for example, compiled guidelines on instrument translation, adaptation, and test use (http://www.intestcom.org/itc_projects.htm/); and the International Society for Quality of Life Research (ISQoL) has a special interest group on translation and cultural adaptation of instruments (<http://www.isoqol.org/>). The European Commission has developed guidelines for health research (Tafforeau, Lopez Cobo, Tolonen, Scheidt-Nave, Tinto, 2005). The International Standards Organization (ISO) has developed the ISO Standard 20252 on Market, Opinion, and Social Research (ISO, 2006). One of the purposes with this global standard is to enhance comparability in international surveys. We already mentioned the cooperation between Eurostat and national agencies on the European Statistical System (http://epp.eurostat.ec.portal/page/portal/about_eurostat/european_framework/ESS/).

New technologies are increasingly being applied to meet the challenges of conducting surveys in remote or inhospitable locations: laptops with extended batteries, “smart” hand-held phones and personal digital assistants (PDAs) that allow transmission of e-mail and data, phones with built-in global positioning systems (GPS), pinpointing an interviewer’s location at all times, digital recorders the size of thumb drives, and geographic information systems (GIS) combined with aerial photography that facilitate sampling in remote regions, to name but a few. It is easy to envision a future when these technologies become affordable and can be used much more widely for quality monitoring in cross-national research.

Both the software tools and the related strategies for analysis have also changed radically for both testing and substantive applications. Statistical applications and models such as Item Response Theory (IRT) and Differential Item Functioning (DIF) have gained popularity as tests for bias, as have, in some instances, Multitrait Multimethod (MTMM) models. The increased availability of courses in instruction, also online, makes it easier for researchers to gain expertise in the new and increasingly sophisticated software and in analytical techniques.

Documentation strategies, tools, and expectations have greatly advanced. One needs only to compare the half-page study reports for the ISSP in the late 1980s with the web-based study monitoring report now required to recognize that a sea change in requirements and transparency is underway. Proprietary and open access databanks help improve consistency within surveys across versions and speed up instrument production, even if the benchmarks for question or translation quality remain to be addressed.

The improved access to data—which itself tends to be better documented than before—is also resulting in a generation of primary and secondary analysts who are better equipped, have plentiful data, and have very different needs and expectations about data quality, analysis, and documentation than researchers of even a decade ago.

Critical mass can make an important difference; the current volume serves as one example: In 2002, regular attendees at cross-cultural survey methods symposia held through the 1990s in ZUMA, Mannheim, Germany, decided to form an annual workshop on “comparative survey design and implementation.” This is

now the International Workshop on Comparative Survey Design and Implementation (CSDI; <http://www.csdiworkshop.org/>). CSDI's organizing committee was, in turn, responsible for organizing the 2008 international conference on Multinational, Multicultural and Multiregional Survey Methods referred to throughout this volume as "3MC" (<http://www.3mc2008.de/>) and were also the prime movers for this present volume. Moreover, work groups at CSDI were the primary contributors to the University of Michigan and University of Nebraska CSDI initiative on cross-cultural survey guidelines mentioned earlier (<http://www.ccsq.isr.umich.edu/>). Finally, although the survey landscape has changed radically in recent years (see Table 1.1), readers not familiar with vintage literature will find much of benefit there and an annotated bibliography is under construction at CSDI (<http://www.csdiworkshop.org/>).

Table 1.1 outlines some of the major developments that have changed or are changing how comparative survey research is conceptualized and undertaken. The abbreviations used in the table are provided at the end of the chapter.

Some of these changes are a natural consequence of developments in the general survey research field. As more modes become available, for example, comparative research avails itself of them as best possible (see Chapter 15, this volume). Other changes are a consequence of the growth in large-scale multipopulation surveys on high-stake research (health, education, policy planning data). The need to address the organizational and quality needs of such surveys has in part been accompanied by funding to allow for more than make-do, ad hoc solutions. Developments there can in turn serve as models for other projects. Finally, the increasing numbers of players in this large field of research and the now quite marked efforts to accumulate and share expertise within programs and across programs are contributing to the creation of a body of information and informed researchers.

1.7 CHALLENGES AND OUTLOOK

3M survey research remains challenging to fund, to organize and monitor, to design, to conduct, and to analyze adequately than research conducted in just one or even two countries. We can mention only a few examples of challenges related to ethical requirements by way of illustration. For example, countries vary widely in official permissions and requirements, as well as in informal rules and customs pertaining to data collection and data access. Heath, Fisher, and Smith (2005) note that North Korea and Myanmar officially prohibited survey research (at the time of reporting), while other countries severely restricted data collection on certain topics or allowed collection but restricted the publication of results (e.g., Iran).

Regulations pertaining to informed consent also vary greatly. American Institutional Review Boards (IRBs) stipulate conditions to be met to ensure respondent consent is both informed and documented. IRB specifications of this particular kind are unusual in parts of Europe, although as Singer (2008) indicates, European regulations on ethical practice can be rigorous. Some European survey practice standards recognize that refusals to participate must be respected, but definitions of what counts as a reluctant or refusing respondent differ